

**IMPLEMENTASI *DATA MINING* DALAM PENERBITAN  
SURAT PENETAPAN TARIF DAN NILAI PABEAN MENGGUNAKAN METODE  
*CLASSIFICATION* PADA DIREKTORAT JENDERAL BEA DAN CUKAI**

**Sri Lestari<sup>1</sup>, Herta Apriani Silaban<sup>2</sup>**

*Program Studi Sistem Informasi, st Ilmu Komputer Cipta Karya Informatika*

Email: <sup>1</sup>sri\_lestari1203@yahoo.com, <sup>2</sup>herta.apriani@gmail.com

---

**ABSTRACT**

*Directorate General of Customs and Excise, which is an agency under the Ministry of Finance which has the duties and strategic functions in carrying out supervision in the field of export imports. Submission of customs documents is self assessment because only company know about the goods imported so needs to be inspected by Customs and Excise Officials. In the event that the customs inspection result is found a wrong notification, the Customs and Excise Officer will make a correction and be stated in a letter called Surat Penetapan Tarif dan Nilai Pabean (SPTNP). Data mining is a term used to describe the discovery of knowledge in a database. Data mining classification method using CRISP-DM research model can be used to provide predictions in the determination of SPTNP. Classification process is done by comparing three classification algorithms, namely Decision Tree, Naive Bayes, and Logistic Regression. The results of the classification process are evaluated using Confusion Matrix and T-Test to get the best accuracy logarithms. This research also uses Correlation Matrix to determine the relationship between factors.*

**Keyword: Data Mining, Classification, CRISP-DM**

---

\* E-mail address: sri\_lestari1203@yahoo.com

## 1. PENDAHULUAN

Direktorat Jenderal Bea dan Cukai, yang merupakan instansi dibawah Kementerian Keuangan Republik Indonesia, juga kini menggunakan sistem komputerisasi baik dalam proses bisnis maupun proses administrasi melalui suatu sistem yang bernama CEISA (*Customs and Excise Integrated System and Automation*). Melalui aplikasi tersebut Pengguna jasa menyampaikan informasi mengenai barang yang diimpornya antara lain berupa jumlah, jenis, spesifikasi barang yang diimpornya termasuk harga barang untuk perhitungan bea masuk. Selanjutnya pengguna jasa menyetorkan pungutan impor ke kas negara melalui bank devisa persepsi. Sistem penyampaian dokumen ini disebut sebagai *self assessment*. Hal ini karena pengguna jasa yang lebih tahu atas barang yang diimpornya. Oleh karena bersifat *self assessment*, penyampaian pemberitahuan impor barang dilakukan pemeriksaan oleh Pejabat Bea dan Cukai. Hal ini untuk memastikan pemberitahuan yang disampaikan oleh pengguna jasa sudah benar dan memenuhi syarat.

Seiring dengan meningkatnya penggunaan komputer, maka penyimpanan dokumen secara digital juga berkembang pesat. Kita kebanjiran data tetapi miskin akan pengetahuan. *Data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. "*Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar" (Turban, dkk. 2005:263). "*Classification* adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui" (Han dan Kamber, 2012:327). Teknik klasifikasi dapat digunakan karena mampu memetakan (mengklasifikasi) sebuah unsur (item) data ke dalam salah satu dari beberapa kelas yang sudah didefinisikan.

Dalam penelitian ini digunakan metode klasifikasi dengan membandingkan tiga algoritma yaitu: *Decision Tree*, *Naive Bayes* dan *Logistic Regression*. Ketiga algoritma tersebut akan dilakukan komparasi untuk dapat diketahui algoritma dengan akurasi terbaik. Serta untuk memperbaiki kinerja (performa) dari masing-masing metode akan digunakan metode *Feature Selection filter (information gain)* dan *Wrapper (Backward Elimination)*.

Tujuan dari penelitian ini adalah mencari korelasi atau hubungan antar atribut-atribut dalam penetapan SPTNP serta mencari faktor/atribut dominan yang paling berpengaruh dalam penetapan SPTNP. Dalam penelitian ini juga dilakukan perbandingan tiga algoritma klasifikasi untuk menentukan penundaan penerbangan dan melakukan perbandingan algoritma *Feature Selection* pada algoritma terbaik. Dengan demikian dapat membantu memberikan prediksi dalam penetapan SPTNP.

Hipotesis merupakan dugaan sementara yang dilakukan oleh peneliti yang akan dibuktikan kebenarannya melalui pengujian. Hipotesis yang diajukan dalam penelitian ini merupakan uraian sementara dari permasalahan yang diajukan, yaitu sebagai berikut:

- H1: Terdapat faktor/atribut dominan yang paling berpengaruh dalam penetapan SPTNP
- H2: Terdapat algoritma klasifikasi data mining dengan akurasi yang paling baik dalam membantu penetapan SPTNP
- H3: Terdapat korelasi atau hubungan antar atribut-atribut dalam penetapan SPTNP
- H4: Implementasi data mining dapat membantu memberikan prediksi dalam penetapan SPTNP

## 2. METODE PENELITIAN

### 2.1. Dataset Penelitian

Dataset yang digunakan dalam penelitian ini adalah data importasi yang dimiliki oleh Direktorat Jenderal Bea dan Cukai yang dikenakan SPTNP.

### 2.2. Dataset Penelitian

### 2.3. Instrumen Penelitian

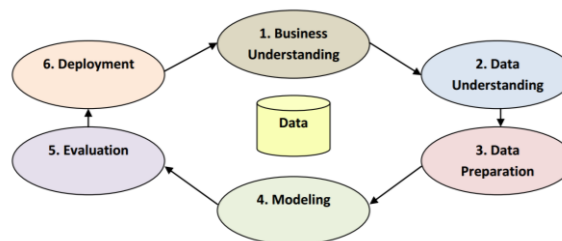
Penelitian ini membutuhkan perangkat keras dan perangkat lunak sebagai penunjang agar dapat berjalan sesuai dengan fungsinya. Kebutuhan tersebut diantaranya:

- a. Kebutuhan Hardware (Perangkat Keras)  
Satu buah *Personal Computer* (PC) untuk melakukan perancangan dan pembangunan data mining, dengan spesifikasi berikut:  
Processor : Intel® Xeon®  
Operating Sistem : Windows 8, 64-bit  
Memory : 16384 MB
- b. Kebutuhan *Software* (Perangkat Lunak)
  - 1) Microsoft Office Excel 2013  
*Software* ini digunakan sebagai media penulisan *datasheet*.
  - 2) RapidMiner Studio 8.2.001  
*Software* yang akan digunakan untuk pengolahan *data mining*, seperti melihat hasil akurasi dari algoritma yang digunakan terhadap dataset yang sedang diteliti

## 2.4. Metode Penelitian

Penelitian ini dilakukan dengan menggunakan metode CRISP-DM. Berikut ini adalah enam tahap metode CRISP-DM (Chapman, 2000:10):

- a. *Business Understanding*
- b. *Data Understanding*
- c. *Data Preparation*
- d. *Modeling*
- e. *Evaluation*
- f. *Deployment*



Gambar 1 Metode CRISP-DM

### 2.4.1. *Business Understanding*

Tahapan ini merupakan fase awal untuk mengetahui masalah yang akan diselesaikan untuk mencapai tujuan yang diinginkan. Kegiatannya antara lain menentukan sasaran atau tujuan bisnis, memahami situasi bisnis, menentukan tujuan data mining dan membuat perencanaan strategi serta jadwal penelitian.

### 2.4.2. *Data Understanding*

Tahap ini diperlukan untuk mempersiapkan data yang akan diolah agar dapat dimodelkan dengan memeriksa apakah data tersebut normal, lengkap dan konsisten sehingga dapat dimodelkan sesuai dengan metode data mining yang akan digunakan. Pada penelitian ini data yang digunakan adalah data primer, yaitu data yang didapatkan langsung dari Direktorat Jenderal Bea dan Cukai dengan ruang lingkup hanya sebatas data importasi yang terkena SPTNP.

### 2.4.3. *Data Preparation*

Tahap ini meliputi semua aktivitas untuk membuat dataset final. Dataset yang dihasilkan seringkali bersifat mentah dan kurang berkualitas, misal terdapat nilai yang hilang, salah input nilai, dan tidak konsisten. Akibatnya perlu dilakukan prapemrosesan data terlebih dahulu. Proses

pembersihan mencakup menghilangkan duplikasi data, mengisi / membuang data yang hilang, memperbaiki data yang tidak konsisten, dan memperbaiki kesalahan ketik.

#### 2.4.4. Modeling

Merupakan fase pemilihan teknik *data mining* dengan menentukan algoritma yang akan digunakan. Dalam tahap ini, berbagai macam teknik pemodelan dipilih dan diterapkan ke dataset yang sudah disiapkan untuk mengatasi kebutuhan bisnis tertentu. Tahap pembuatan model juga mencakup penilaian dan analisa komparatif dari berbagai model yang dibangun. Algoritma klasifikasi yang akan digunakan dalam penelitian ini adalah *Decision Tree*, *Naive Bayes* dan *Logistic Regression*. Ketiga algoritma tersebut akan dilakukan komparasi untuk dapat diketahui algoritma dengan akurasi terbaik. Serta untuk memperbaiki kinerja (performa) dari masing-masing metode akan digunakan metode *Feature Selection filter (information gain)* dan *Wrapper (Backward Elimination)*.

#### 2.4.5. Evaluation

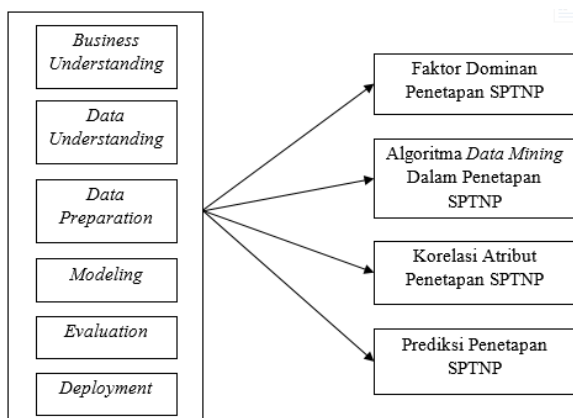
Dalam tahapan ini akan dilakukan validasi serta pengukuran keakuratan hasil yang dicapai oleh model yang telah dibuat. Untuk membandingkan performa kinerja dari ketiga algoritma tersebut, digunakan Uji Beda (*T-Test*). Dengan *T-Test* dapat diketahui akurasi untuk klasifikasi dan perbedaan signifikan dari ketiga algoritma tersebut. Untuk mengetahui hubungan antar faktor atribut digunakan *Correlation Matrix* yang dapat mendeskripsikan bentuk dan kekuatan hubungan antar faktor tersebut.

#### 2.4.6. Deployment

Pembuatan dari model bukanlah akhir dari proyek *data mining*. Meskipun tujuan dari pemodelan adalah untuk meningkatkan pengetahuan dari data, pengetahuan data tersebut perlu dibangun dengan terorganisasi dan dibuat pada satu bentuk yang dapat digunakan oleh pengguna. Tahap ini merupakan tahap implementasi (penyebaran) dari *data mining*.

### 2.5. Kerangka Pemikiran

Peneliti akan menggunakan metode CRISP-DM pada klasifikasi yang dapat membantu penetapan SPTNP. Secara rinci kerangka pemikirannya dapat dilihat pada gambar berikut:



Gambar 2 Kerangka pemikiran

## 3. HASIL DAN PEMBAHASAN

### 3.1. Business Understanding

#### 3.1.1. Motivasi

- Pengisian dokumen kepabeanaan bersifat *self assessment* yang menyebabkan kemungkinan terjadi kesalahan dalam pengisian data.
- Dokumen kepabeanaan impor perlu dilakukan pemeriksaan oleh Pejabat Bea dan Cukai.

- c. Dalam hal terjadi kesalahan dalam pengisian dokumen, maka akan dituangkan dalam surat penetapan tarif dan nilai pabean atas penelitian dokumen impor yang dilakukan oleh Pejabat Bea dan Cukai atau disebut dengan SPTNP.
- d. Pejabat Bea dan Cukai perlu mengetahui faktor-faktor yang berpengaruh dalam penetapan SPTNP sehingga dengan mengetahui hubungan antar faktor tersebut, Pejabat Bea dan Cukai dapat mengambil keputusan yang lebih baik dalam penetapan SPTNP.
- e. Pejabat Bea dan Cukai memilih korelasi sebagai cara untuk model hubungan antar faktor untuk diketahui. Korelasi adalah pengukuran statistik bagaimana kekuatan hubungan antar atribut dalam Dataset.
- f. Pejabat Bea dan Cukai juga perlu untuk mencari metode klasifikasi yang tepat dalam menentukan penetapan SPTNP.
- g. Metode klasifikasi yang tepat dapat membantu Pejabat Bea dan Cukai dalam menentukan penetapan SPTNP dengan baik.

### 3.1.2. Objektif

- a. Untuk mencari hubungan antar faktor yang mempengaruhi penetapan SPTNP.
- b. Untuk mencari metode klasifikasi yang terbaik dengan melakukan perbandingan tiga algoritma klasifikasi.
- c. Untuk meningkatkan performa dari metode klasifikasi dapat dilakukan dengan menggunakan *Feature Selection*.

### 3.2. Data Understanding

Dengan menggunakan sumber data yang didapatkan langsung dari Direktorat Jenderal Bea dan Cukai, dibuat dataset dengan atribut sebagai berikut:

- a. NPWP: Nomor Pokok Wajib Pajak dari perusahaan
  - b. Profil: menunjukkan profil resiko perusahaan
  - c. Status perusahaan: menunjukkan status dari perusahaan (Importir produsen, Importir Umum)
  - d. HS\_CODE (Komoditi): menunjukkan komoditi yang diimpor oleh perusahaan
  - e. Jalur: menunjukkan status penjaluran dokumen impor perusahaan
- Flag SPTNP: merupakan flag yang menunjukkan bahwa dokumen tersebut terkena SPTNP (0=tidak terkena SPTNP; 1=terkena SPTNP)

NPWP	Profil	Status Perusahaan	HS_CODE	Jalur	FL_SPTNP
010	Medium	LAINNYA	85044090	Hijau	0
010	Medium	IP - IMPORTIR PRODUSEN	90278010	Hijau	0
010	Medium	IP - IMPORTIR PRODUSEN	84213990	Hijau	0
010	Medium	IP - IMPORTIR PRODUSEN	85366939	Hijau	0
013	Medium	LAINNYA	84818099	Hijau	0
010	Low	IP - IMPORTIR PRODUSEN	29124990	Hijau	0
010	Low	IP - IMPORTIR PRODUSEN	33011900	Hijau	0
010	Low	IP - IMPORTIR PRODUSEN	29339200	Hijau	0
010	Low	IP - IMPORTIR PRODUSEN	13021400	Hijau	0
010	Low	IP - IMPORTIR PRODUSEN	29124910	Hijau	0
020	Medium	IU - IMPORTIR UMUM	21023000	Hijau	0
018	Medium	MITA	90299020	Mita Non Prioritas	0
020	Medium	LAINNYA	85122099	Hijau	0
020	Medium	LAINNYA	87088016	Hijau	0
020	Medium	LAINNYA	84099149	Hijau	0
020	Medium	LAINNYA	85129020	Hijau	0
020	Medium	LAINNYA	85361099	Hijau	0
020	Medium	LAINNYA	84839019	Hijau	0
010	Medium	MITA	33049990	Mita Prioritas	0
010	Medium	MITA	40118011	Mita Prioritas	0
018	Medium	IP - IMPORTIR PRODUSEN	85129020	Hijau	0
010	Medium	MITA	74112100	Mita Prioritas	0
010	Medium	MITA	74101100	Mita Prioritas	0
010	Medium	MITA	85322900	Mita Prioritas	0
010	Medium	MITA	39094010	Mita Prioritas	0
010	Low	MITA	85364140	Mita Prioritas	0
013	Medium	AEO	39100020	Mita Prioritas	0
013	Medium	AEO	39069099	Mita Prioritas	0
010	Low	IU - IMPORTIR UMUM	28121100	Hijau	0
017	Medium	LAINNYA	38123900	Hijau	0
020	High	LAINNYA	17026020	Hijau	0
020	High	LAINNYA	17026020	Merah	0
010	High	LAINNYA	39074000	Hijau	0
020	Low	IP - IMPORTIR PRODUSEN	39204900	Merah	0

Gambar 3 Dataset Importasi SPTNP

### 3.3. Data Preparation

Setelah dilakukan evaluasi terhadap kualitas datanya ternyata terdapat missing value pada datasetnya yaitu pada atribut NPWP. Apabila diperhatikan dalam statistik akan tampak seperti gambar 3.



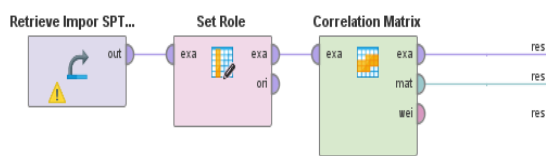
Name	Type	Missing	Statistics
NPWP	Integer	11	Min: 10000347051000, Max: 842534364503000, Average: 34705127311379.960
Profil	Polynomial	0	Low (66), Medium (4520), Very High (3534), [2 more]
Status Perusahaan	Polynomial	0	PTM - PERTAMINA (0), LABOYA (4888), LANOYA (4888), NI - IMPORTIR UMBU (4155), [4 more]
HS_CODE	Integer	0	Min: 90019000, Max: 9004646.788
Jalur	Polynomial	0	Mta Non Prioritas (1), Hibu (4162), Hibu (4162), Merah (4009), [4 more]
PL_SPTNP	Integer	0	Min: 0, Max: 1, Average: 0.237

Gambar 4 Statistik Dataset Importasi SPTNP

### 3.4. Modeling

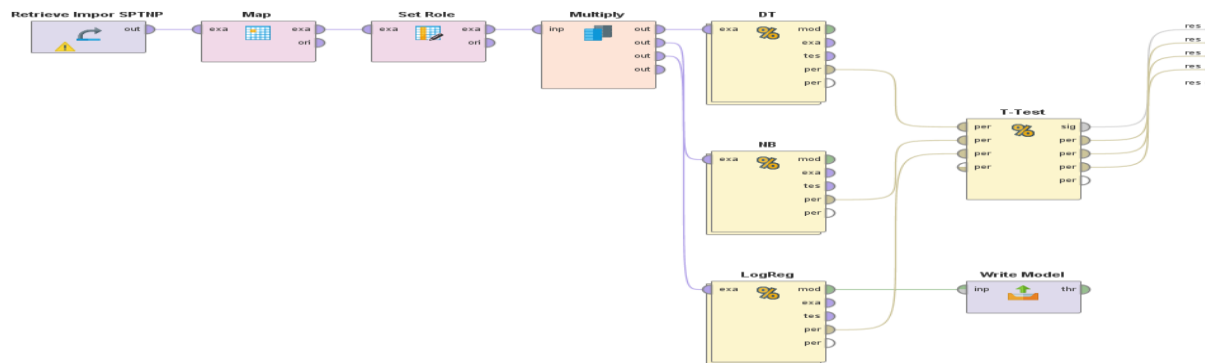
Merupakan fase pemilihan teknik *data mining* dengan menentukan algoritma yang akan digunakan. *Tools* yang digunakan adalah RapidMiner Studio versi 8.2.001.

Untuk mengetahui korelasi antar atribut digunakan *Correlation Matrix* yang dapat mendeskripsikan bentuk dan kekuatan antar atribut tersebut.



Gambar 5 Model Correlation Matrix

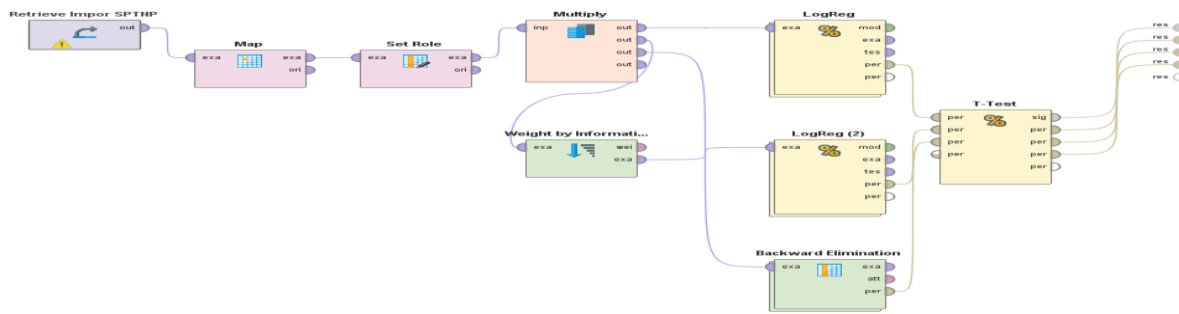
Penelitian ini menggunakan metode *Cross Validation* untuk membandingkan tiga algoritma klasifikasi yaitu *Decision Tree*, *Naive Bayes* dan *Logistic Regression*.



Gambar 6 Komparasi Algoritma Klasifikasi

Dan menggunakan uji beda (*T-Test*) untuk membandingkan kinerja (performa) dari ketiga algoritma tersebut. Dengan uji beda (*T-Test*) dapat diketahui akurasi untuk klasifikasi dan perbedaan signifikan dari ketiga algoritma tersebut untuk dianalisa sehingga dapat diketahui yang terbaik.

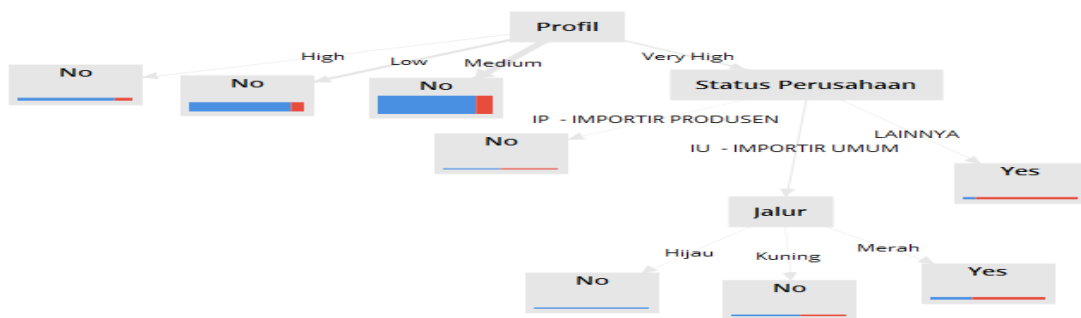
Untuk memperbaiki kinerja (performa) dari algoritma yang telah dipilih (algoritma dengan akurasi terbaik), dapat digunakan metode *Feature Selection*, yaitu *Information Gain* dan *Backward Elimination*.



Gambar 7 Model Komparasi *Feature Selection*

3.5. *Evaluation*

a. Pohon Keputusan



Gambar 8 Pohon Keputusan Penetapan SPTNP

Pola/pengetahuan yang didapatkan dari pohon keputusan pada Gambar 8 adalah:

1. Apabila profil perusahaan *Very High* dan status perusahaan lainnya, maka dominan terkena SPTNP
2. Apabila profil perusahaan *Very High* dengan status perusahaan Importir Umum dan jalur dokumen merah, maka dominan terkena SPTNP
3. Apabila profil perusahaan *Very High* dengan status perusahaan Importir Umum dan jalur dokumen hijau, maka tidak terkena SPTNP

Tree

```

Profil = High: No {No=11, Yes=2}
Profil = Low: No {No=47, Yes=6}
Profil = Medium: No {No=94, Yes=16}
Profil = Very High
| Status Perusahaan = IP - IMPORTIR PRODUSEN: No {No=1, Yes=1}
| Status Perusahaan = IU - IMPORTIR UMUM
| | Jalur = Hijau: No {No=3, Yes=0}
| | Jalur = Kuning: No {No=3, Yes=2}
| | Jalur = Merah: Yes {No=4, Yes=7}
| Status Perusahaan = LAINNYA: Yes {No=1, Yes=8}
    
```

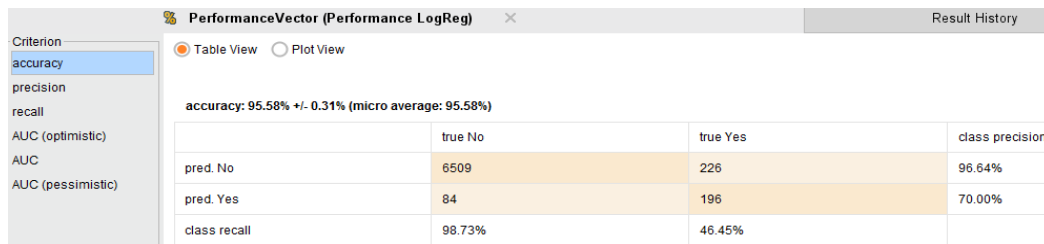
Gambar 9 *Rule* Pohon Keputusan Penetapan SPTNP

b. *Accuracy*

Berdasarkan proses *Cross Validation*, diperoleh hasil bahwa algoritma yang paling baik digunakan untuk dataset penentuan SPTNP adalah algoritma *Logistic Regression* karena memiliki tingkat akurasi terbaik, yaitu 95.58%. Diurutan kedua adalah algoritma *Decision Tree* dengan tingkat akurasi 93.98%.

Tabel 1 Hasil Akurasi Algoritma Klasifikasi

Algoritma	Accuracy	AUC
Decision Tree	93.98%	0.500
Naive Bayes	91.33%	0.902
Logistic Regression	95.58%	0.888



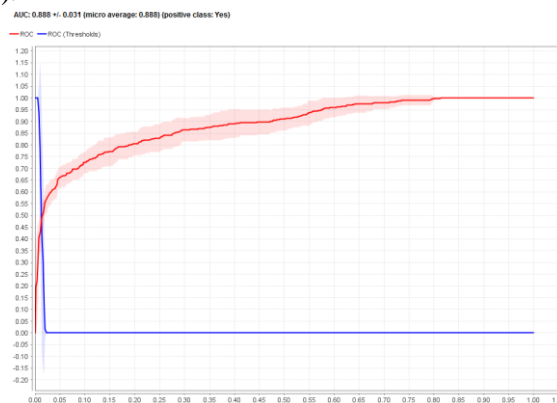
Gambar 10 Confusion Matrix Algoritma Logistic Regression

A	B	C	D
	0.940 +/- 0.001	0.913 +/- 0.008	0.956 +/- 0.003
0.940 +/- 0.001		0.000	0.000
0.913 +/- 0.008			0.000
0.956 +/- 0.003			

Gambar 11 Hasil Uji Beda (T-Test) Algoritma Klasifikasi

c. ROC

Kurva Receiver Operating Characteristic (ROC) digunakan untuk mengekspresikan data Confusion Matrix. Garis horizontal mewakili nilai False Positives (FP) dan garis vertikal mewakili nilai True Positives (TP).



Gambar 12 Grafik ROC Algoritma Logistic Regression

Kategori Klasifikasi AUC:

1. 0.90 - 1.00 = *excellent classification*
  2. 0.80 - 0.90 = *good classification*
  3. 0.70 - 0.80 = *fair classification*
  4. 0.60 - 0.70 = *poor classification*
  5. 0.50 - 0.60 = *failure*
- (Gorunescu, 2011:325)



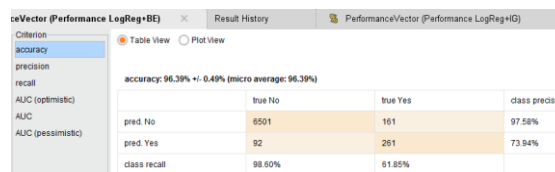
Dari kurva ROC Algoritma *Logistic Regression* memiliki AUC sebesar 0.888 ini berarti termasuk dalam kategori *good classification*.

d. *Feature Selection*

Dari hasil komparasi *Feature Selection*, diketahui bahwa metode yang paling baik untuk meningkatkan akurasi algoritma adalah *Backward Elimination*. Dengan *Feature Selection* ini performanya lebih baik dari sebelumnya karena akurasi meningkat 0.81 dari 95.58% menjadi 96.39% sehingga model terbaik yang digunakan adalah *Logistic Regression + Backward Elimination*. Hasil peningkatan akurasi dapat dilihat pada tabel berikut:

Tabel 2 Hasil Komparasi *Feature Selection*

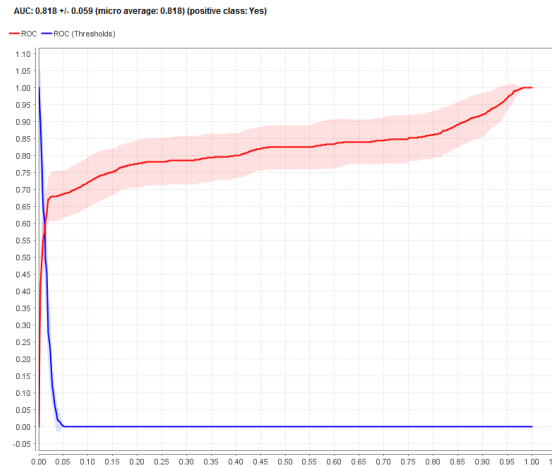
Algoritma	Accurac	AUC
<i>Logistic Regression</i> (LR)	95.58%	0.888
LR + <i>Information Gain</i>	95.68%	0.889
LR + <i>Backward Elimination</i>	96.39%	0.888



Gambar 13 *Confusion Matrix* Algoritma LR + *Backward Elimination*

A	B	C	D
	0.956 +/- 0.006	0.957 +/- 0.009	0.964 +/- 0.005
0.956 +/- 0.006		0.837	0.006
0.957 +/- 0.009			0.039
0.964 +/- 0.005			

Gambar 14 Hasil Uji Beda (T-Test) *Feature Selection*



Gambar 15 Grafik ROC Algoritma LR + *Backward Elimination*

Dari kurva ROC dapat diketahui Algoritma LR + *Backward Elimination* memiliki AUC sebesar 0.818 ini berarti termasuk dalam kategori *good classification*.

e. Uji Coba Model

Row No.	predicto...	confidenc...	confidenc...	FL_SPTNP	NP...	Profil	Status Pers...	HS_CODE	Jalur
25	Yes	0.000	1.000	Yes	013..	Low	IP - IMPOR...	20059990	Hijau
31	Yes	0.000	1.000	Yes	312..	High	LAANNYA	48193000	Kuning
39	Yes	0.024	0.976	Yes	030..	Medium	IP - IMPOR...	38140000	Kuning
40	Yes	0	1	Yes	031..	Very High	LAANNYA	84143090	Merah
48	Yes	0	1	Yes	031..	Very High	LAANNYA	94013000	Merah
54	Yes	0	1	Yes	736..	Very High	IJ - IMPOR...	84798939	Kuning
57	Yes	0.000	1.000	Yes	317..	Medium	LAANNYA	84483200	Merah
58	Yes	0.000	1.000	Yes	020..	Low	IP - IMPOR...	48010011	Hijau
63	Yes	0.000	1.000	Yes	031..	Very High	LAANNYA	40169190	Merah
66	Yes	0.000	1.000	Yes	030..	Very High	LAANNYA	68052000	Kuning
67	Yes	0.073	0.927	Yes	010..	Medium	LAANNYA	84149091	Hijau
73	Yes	0.367	0.633	Yes	210..	Medium	IJ - IMPOR...	02022000	Kuning
82	Yes	0.000	1.000	Yes	726..	Medium	IJ - IMPOR...	68029110	Merah
89	Yes	0.480	0.520	Yes	315..	Very High	IJ - IMPOR...	85444229	Merah
94	Yes	0.000	1.000	Yes	018..	Medium	IP - IMPOR...	25151220	Merah
95	Yes	0	1	Yes	030..	Very High	IJ - IMPOR...	39249090	Kuning
97	Yes	0.001	0.999	Yes	029..	Medium	IJ - IMPOR...	90021900	Kuning
103	Yes	0.000	1.000	Yes	754..	Very High	LAANNYA	85393110	Merah
106	Yes	0.000	1.000	Yes	013..	Medium	LAANNYA	49019990	Hijau
107	Yes	0.000	1.000	Yes	020..	Very High	IJ - IMPOR...	40169500	Merah
108	Yes	0.000	1.000	Yes	017..	Low	LAANNYA	29181900	Hijau

Gambar 16 Hasil Uji Coba Algoritma Dengan Data *Testing*

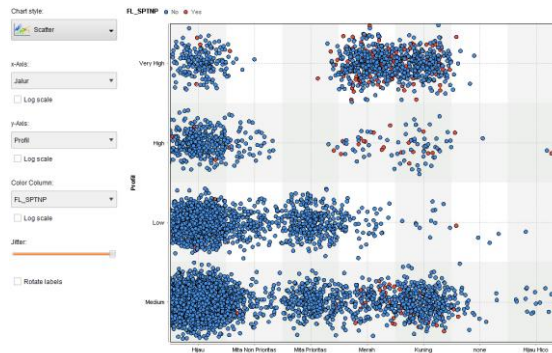
f. *Correlation Matrix*

Dari hasil pemodelan *Correlation Matrix*, maka diperoleh:

1. Atribut (faktor) yang paling signifikan berpengaruh pada penetapan SPTNP adalah Profil Perusahaan (hubungan positif)
2. Atribut (faktor) kedua yang paling berpengaruh adalah Jalur Dokumen (hubungan positif)
3. Atribut (faktor) Hs Code / Komoditi tidak terlalu berpengaruh pada penetapan SPTNP
4. Atribut (faktor) Status Perusahaan boleh dikatakan tidak berpengaruh pada penetapan SPTNP

Selain itu dapat diketahui juga kekuatan hubungannya. Apabila semakin besar nilai *correlation*nya maka semakin kuat/banyak hubungannya. Begitu juga sebaliknya semakin kecil nilai *correlation*nya maka semakin lemah/sedikit hubungannya. Contohnya adalah hubungan antara profil perusahaan dengan komoditi memiliki hubungan yang sangat kuat karena resiko profil perusahaan sangat tergantung pada komoditi yang diimpor. Sedangkan untuk yang hubungannya lemah/sedikit contohnya adalah hubungan antara status perusahaan dengan komoditi (Hs Code) karena status perusahaan tidak tergantung pada komoditi yang diimpor. Sedangkan yang nilai *correlation*nya sangat kecil bisa dikatakan tidak berhubungan.

Attributes	Profil	Status Perusahaan	HS_CODE	Jalur	FL_SPTNP
Profil	1	-0.092	0.146	0.068	0.265
Status Perusahaan	-0.092	1	0.015	-0.066	0.024
HS_CODE	0.146	0.015	1	0.023	0.061
Jalur	0.068	-0.066	0.023	1	0.093
FL_SPTNP	0.265	0.024	0.061	0.093	1

Gambar 17 Hasil *Correlation Matrix*Gambar 18 *Scatter* Importasi SPTNP Berdasarkan Profil dan Jalur

Dari Gambar 18 dapat diketahui bahwa dominan importasi yang terkena SPTNP adalah impor oleh perusahaan profil dengan *very high* dan jalur dokumen merah dan kuning. Dan juga perusahaan dengan profil medium dan jalur dokumen kuning. Sedangkan pada perusahaan dengan jalur MITA Prioritas tidak terdapat dokumen SPTNP.

### 3.6. Deployment

Telah dihasilkan suatu informasi, dan pola pengetahuan baru dalam proses *data mining*. Pola pengetahuan tersebut didapat dari metode Korelasi, Klasifikasi, dan *Feature Selection* untuk menentukan SPTNP berdasarkan dataset importasi SPTNP yang terdapat pada Direktorat Jendral Bea dan Cukai. Untuk atribut yang tidak terlalu berpengaruh dalam metode klasifikasi tersebut dapat dihilangkan seperti NPWP.

Klasifikasi penetapan SPTNP sangat tergantung pada penjaluran dokumen akibat status profil perusahaan. Keakuratan algoritma klasifikasi dapat ditingkatkan dengan menggunakan *Backward Elimination* sehingga dapat menghasilkan keputusan klasifikasi yang lebih akurat. Pengetahuan yang diperoleh dapat digunakan sebagai dasar dalam mengambil keputusan untuk menentukan penetapan SPTNP sehingga pejabat pemeriksa dapat melakukan mitigasi untuk mengurangi resiko kesalahan penetapan SPTNP dan dapat memberikan prediksi penetapan SPTNP terhadap dokumen impor.

## 4. KESIMPULAN

- a. Faktor dominan dalam penetapan SPTNP.
  - 1) Faktor yang paling berpengaruh dalam penetapan SPTNP adalah profil perusahaan
  - 2) Faktor yang tidak terlalu berpengaruh dalam penetapan SPTNP adalah status perusahaan
  - 3) Semakin tinggi resiko dari profil perusahaan, semakin tinggi juga status penjaluran dokumennya dan semakin berpengaruh terhadap penetapan SPTNP
- b. Algoritma Klasifikasi
  - 1) Algoritma dengan akurasi terbaik adalah *Logistic Regression* dengan akurasi 76.35% dan AUC 0.745 dan tidak memiliki perbedaan yang cukup signifikan
  - 2) Metode *Feature Selection* yang terbaik untuk Algoritma *Logistic Regression* pada penelitian ini adalah *Backward Elimination*

- 3) Metode dengan model *Logistic Regression + Backward Elimination* memiliki tingkat akurasi yang tinggi sehingga metode ini dapat digunakan sebagai rekomendasi dalam membantu mengambil keputusan yang tepat untuk menetapkan SPTNP
- c. Hubungan antar faktor / atribut
- 1) Hubungan antara flag SPTNP dengan profil perusahaan memiliki hubungan positif dan sangat kuat karena penetapan SPTNP sangat dipengaruhi oleh profil perusahaan
  - 2) Hubungan antara profil perusahaan dengan komoditi memiliki hubungan positif karena resiko profil perusahaan sangat tergantung pada komoditi yang diimpor.
  - 3) Hubungan antara profil perusahaan dengan status importir memiliki hubungan negatif karena semakin tinggi profil perusahaan maka status perusahaan akan semakin rendah
- d. Prediksi SPTNP
- Metode CRISP-DM pada klasifikasi dapat digunakan untuk mendapatkan algoritma terbaik dalam memberikan prediksi penetapan SPTNP. 99% hasil prediksi sesuai dengan data yang sebenarnya.
- e. Saran
- Untuk memperoleh hasil prediksi penetapan SPTNP yang lebih baik dapat dilakukan penelitian dengan menggunakan algoritma lain yang lebih baru dan lebih baik. Selanjutnya untuk lebih meningkatkan performa dari algoritma tersebut dapat dilakukan dengan menggunakan metode *Feature Selection* atau *Feature Extraction* lain yang dapat meningkatkan performa algoritma menjadi lebih cepat dan lebih akurat.

## 5. REFERENSI

- Chapman, Pete, dkk. 2000. *CRISP-DM v.1.0 Step-by-step data mining guide*. SPSS Inc.
- Gorunescu, F. 2011. *Data Mining: Concepts, Models and Techniques*. Berlin: Springer-Verlag Berlin Heidelberg.
- Han, J, Kamber, M, & Pei, J. 2012. *Data Mining: Concept and Techniques, Third Edition*. Waltham: Morgan Kaufmann Publishers.
- Turban, E., dkk. 2005. *Decision Support Systems and Intelligent Systems*. Yogyakarta: Andi Offset.