

# PENGGUNAAN METODE SUPPORT VECTOR MACHINE UNTUK MENGLASIFIKASI DAN MEMPREDIKSI ANGKUTAN UDARA JENIS PENERBANGAN DOMESTIK DAN PENERBANGAN INTERNASIONAL DI BANDA ACEH

Sayed Fachrurrazi<sup>1</sup>, Burhanuddin<sup>2</sup>  
Sistem Informasi Universitas Malikussaleh  
Jl. Cot Tgk Nie-Reulet, Aceh Utara, 141 Indonesia  
Email: sayedfachrurrazi@gmail.com, burhanuddin.sipil@gmail.com

## ABSTRAK

Penelitian ini menyajikan analisis performansi Support Vector Machine (SVM) dengan 11 variabel bebas dan 1 variabel terikat. Metode SVM dengan data training (75%) dan data testing (25%) yang digunakan pada pengklasifikasian data Penerbangan domestic dan data penerbangan internasional untuk menemukan hyperplane terbaik yang memisahkan dua buah kelas. Hasilnya terdapat 4 support vector memberikan informasi yang dibutuhkan untuk menyakinkan bahwa metode SVM bias sebagai classifier dan dapat memprediksi keakuratan model dengan menggunakan kurva *Receiver Operating Characteristic (ROC)* untuk melihat akurasi model terbaik. mencapai 84,31%.

**Kata Kunci:** Klasifikasi, Metode Support Vector Machine (SVM), Receiver Operating Characteristic (ROC)

## 1. PENDAHULUAN

Setelah di launchingnya program tahun kunjungan wisata di tahun 2011 ini, Pemerintah daerah kota banda aceh beserta masyarakat sangat gencar mencanangkan *Visit Banda Aceh 2011* dengan motto "Bandar Wisata Islami Indonesia", *A place blessed with natural beauty and as a spiritual gateway*.

Salah satu bidang yang berkembang sangat pesat sejalan dengan industri pariwisata dan perkembangan teknologi informatika saat ini adalah *Data Mining* (DM). *Data Mining* disebut juga *knowledge discovery in database* yaitu kegiatan yang meliputi pengumpulan, penggunaan data historis untuk mendapat keteraturan, pola atau hubungan dalam suatu data. Keluaran yang dihasilkan dapat digunakan untuk pengambilan keputusan atau perencanaan dimasa yang akan datang.

## 2. TINJAUAN PUSTAKA

*Support Vector Machine* (SVM) adalah salah satu bagian dari *Data Mining* yang akhir-akhir ini mendapat perhatian dalam hal untuk pengklasifikasian (Nugroho, 2003)

Dengan memanfaatkan data induk jenis penerbangan baik lokal maupun internasional yang berkunjung ke aceh khususnya banda aceh selama periode bulan januari 2010 hingga bulan maret tahun 2011 maka akan dapat diklasifikasi dan diprediksi jumlah kedatangan wisatawan yang berkunjung ke banda aceh dengan memanfaatkan data dari dinas pariwisata dan kebudayaan, data dari kantor imigrasi wilayah I, data dari bandara internasional sultan iskandar muda blang bintang, data dari biro perjalanan wisata, data dari hotel tempat wisatawan menginap, dan data lainnya yang mendukung.

Perkembangan informasi online yang pesat saat ini menjadikan klasifikasi sebagai suatu teknik utama dalam menangani dan mengelola data teks. **Nugroho, 2003** menerangkan bahwa klasifikasi dokumen merupakan suatu proses pengklasifikasian data teks ke dalam satu atau lebih kategori yang telah didefinisikan sebelumnya atau ke dalam kelompok yang sama. **Brefeld Ulf, 2005** menerangkan bahwa salah satu metode klasifikasi adalah *Support Vector Machine* (SVM). SVM akan mencari support vector terbaik yang memisahkan dua buah class dengan margin terbesar. SVM secara konseptual merupakan classifier yang bersifat linier, tetapi SVM dapat dimodifikasi dengan menggunakan kernel (fungsi yang memudahkan proses pengklasifikasian data) sehingga dapat menyelesaikan permasalahan yang bersifat tidak linier (*non linier*) (Sun Aixin, 2002). SVM merupakan metode baru, namun dapat memberikan performansi yang lebih baik bila dibandingkan metode-metode

pengklasifikasian lainnya seperti *Naive Bayesian*, *Rocchio*, *C4.5* dan *k-N* (Karatzoglou Alexandros, 2006)

### 3. METODOLOGI PENELITIAN

Penelitian ini dilaksanakan di Bandara Sultan Iskandar Muda Blang Bintang Banda Aceh periode bulan Januari hingga Desember 2017 dengan menggunakan data primer dan data sekunder yang diperoleh secara langsung dari pihak-pihak yang berkepentingan. Juga dikuatkan dengan data-data jumlah wisatawan yang berkunjung ke provinsi aceh baik wisatawan lokal maupun wisatawan mancanegara dari Dinas Pariwisata Provinsi Aceh. Serta dilakukan wawancara dengan pihak-pihak yang terkait dengan industri pariwisata baik pihak manajemen hotel, losmen, asosiasi pariwisata dan pihak pendukung lainnya.

### 4. HASIL DAN PEMBAHASAN

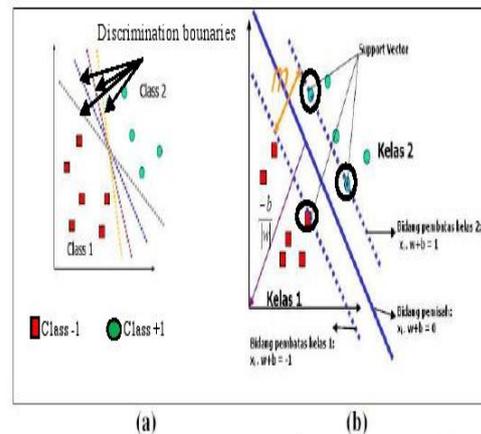
#### 4.1 Analisis Training Set Menggunakan Binary SVM Light

Untuk menentukan training set dengan Software Open Source Perangkat Lunak R, maka dilakukan langkah awal dengan menginstall packages, yaitu :

```
> install.packages ('RWeka')
package 'RWeka' successfully unpacked and MD5 sums checked
The downloade packages are in
C : \Documents and Settings\sayed\Local
> library (RWeka)
> data.penerbangan
Tampilan data.penerbangan (pada lampiran)
> set.seed (123)
> N = dim (data.penerbangan) [1
> N
[1] 30
```

Berdasarkan data dari lampiran, `dim (data.penerbangan) [1]`, Jumlah data angkutan udara secara total adalah angkutan  $N=30$ . Konsep SVM sebagai usaha mencari hyperplane terbaik yang memisahkan dua buah kelas dengan margin terbesar. Beberapa pola yang merupakan anggota dari dua buah kelas adalah positif (+1) dan negatif (-1). Pada proses pembelajaran dalam masalah klasifikasi diterjemahkan sebagai upaya menemukan *hyperplane* yang memisahkan antara kedua kelompok tersebut. Berbagai alternatif garis pemisah (*discrimination boundaries*) dan *hyperplane* terbaik antara kedua kelas dapat ditemukan dengan mengukur margin

*hyperplane* tersebut dan mencari titik maksimalnya. Margin adalah jarak antara *hyperplane* dengan data terdekat dari masing-masing kelas.



**Gambar 4.1 SVM berusaha menemukan hyperplane yang terbaik yang memisahkan kedua class positif dan negatif**

Bagian data training set yang paling dekat dengan hyperplane ini disebut *support vector*. Pada tahap untuk membangun model pada penelitian ini dengan training.penerbangan pada SVM terdapat 4 support vector pada garis hyperplane. Model yang dibangun memiliki 2 model untuk penerbangan domestik dan 2 model untuk penerbangan internasional. Garis tebal pada gambar 4.1 (b) menunjukkan *hyperplane* terbaik yang terletak tepat ditengah-tengah kedua kelas, sedangkan titik kotak dan lingkaran yang berada dalam lingkaran hitam adalah *support vector*. Dalam penelitian ini model dibangun dengan menggunakan kernel yang telah disediakan oleh paket SVM<sup>light</sup>. Kernel yang terdapat pada paket SVM<sup>light</sup> ada 4 (empat) yaitu *linear*, *polynomial*, *radial*, dan *sigmoid*. Pada penelitian ini hanya digunakan 1 (satu) kernel yaitu *linear*.

Dari data pada lampiran listing program dapat dilihat bahwa keakuratan analisis training data mencapai 0,8431417 atau 84,31 % keakuratan data yaitu terdapat 12 variabel termasuk 1 variabel Y yang mempunyai 2 label yaitu domestik dan internasional yang ada pada model. Pada data training..penerbangan yang dipergunakan untuk membangun model terdapat 22 training data (75%) yang lebih banyak dari data testing

dan 11 attribute sedangkan pada data testing.penerbangan terdapat 8 data testing (25%) dengan 11 Attribute juga.

#### 4.2 Analisis Training dan Testing Set Menggunakan ROC

Untuk mengklasifikasikan data sehingga mendapatkan model terbaik dari SVM maka perlu dilakukan langkah-langkah sebagai berikut :

Data yang telah diinputkan dibagi dalam 2 bagian yakni training set dan testing set. Adapun pembagian untuk data training set adalah sebesar 75 % dari data awal sedangkan sisanya untuk data testing set sebesar 25% yang diambil secara random.

$x = \{x_0, x_1, x_2, x_3, \dots, x_n\}$  : sampel *training*

$y = \{y_0, y_1, \dots, y_m\} \in \{\pm 1\}$  : label data *training*

Sebelum data diolah dengan menggunakan metode *support vector machine*, maka dilakukan preprocessing data terlebih dahulu. Yakni dengan menormalisasikannya menggunakan library (RWeka) dari perangkat lunak R. Hal ini dilakukan untuk mendapatkan hasil analisis yang lebih akurat tanpa merubah informasi yang dikandung didalamnya. Setelah melakukan tahap pembelajaran dan dihasilkan model klasifikasi, langkah selanjutnya adalah melakukan pengklasifikasian terhadap data testing. Evaluasi hasil didasarkan pada tingkat keakurasian data yang terklasifikasi dengan benar. Semakin tinggi nilai akurasi, maka akan menunjukkan tingkat keberhasilan yang pula dari system dalam mengklasifikasikan data ke dalam kategori-kategorinya.

Model pada penelitian ini juga dibangun menggunakan SVM Light yang terdapat pada perangkat lunak R.Model yang didapat diuji keakuratannya dengan menggunakan Receiver Operating Characteristics (ROC) yang terdapat didalam paket versi coTools 1.10 pada R.Model yang akurat adalah model yang memiliki nilai Area Under Curve (AUC) mendekati satu dan nilai kurva ROC > 0,5.

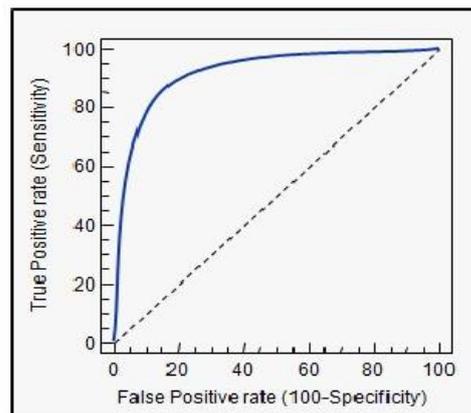
AUC (Area Under Curve) adalah luas daerah di bawah kurva ROC. Apabila nilainya mendekati satu, maka model yang didapat lebih akurat. Berdasarkan gambar di atas maka dapat dilihat karakteristik dari AUC adalah sebagai berikut :

- Area maksimum adalah 1
- Jika ROC = 0,5 maka model yang dihasilkan belum terlihat optimal

Sedangkan jika ROC > 0,5 maka model yang dihasilkan akan lebih baik.

Yang menjadi variabel-variabel dalam penelitian ini adalah variabel bebas (Y) dan variabel terikat (X), yaitu :

- Y = Jenis Penerbangan (domestik dan internasional)
- X1 = Pesawat yang datang
- X2 = Pesawat yang berangkat
- X3 = Pesawat lokal
- X4 = Penumpang yang datang
- X5 = Penumpang yang berangkat
- X6 = Bagasi dari kedatangan pesawat
- X7 = Bagasi dari pesawat yang berangkat
- X8 = Kargo dari pesawat yang datang
- X9 = Kargo dari pesawat yang berangkat
- X10 = Pos dari pesawat yang datang
- X11 = Pos dari pesawat yang berangkat



**Gambar 4.2. Kurva ROC dengan luas area AUC terbaik**

### 4.3 Interpretasi Kurva ROC

Gambar 4.2 adalah kurva ROC yang memiliki luas area AUC hanya sebesar 84,31%. Artinya model yang terbaik yaitu berdasarkan kurva ROC yang tertinggi sebesar 84,31% dengan parameternya kernel linear. Model yang didapat telah lebih bagus sebesar 84,31%.

Pada data.penerbangan, model yang dihasilkan oleh kernel linear berdasarkan kurva ROC dari kategori diatas adalah diatas 50%. Akurasi

model tertinggi berdasarkan kurva ROC dihasilkan oleh kernel linear adalah sebesar 84,31% artinya model yang dihasilkan memiliki keakuratan hingga 84,31% dengan support vector sebanyak 4 support vector pada garis hyperplane yang terdapat pada satu bidang untuk kelas penerbangan domestik dan kelas penerbangan internasional, support vector yang dihasilkan oleh model pengklasifikasian.

Pada tahapan pengklasifikasian dokumen jenis penerbangan angkutan udara adalah suatu tahap pemberian kategori yang telah didefinisikan kepada dokumen angkutan udara yang belum memiliki kategori. Mengklasifikasikan dokumen jenis penerbangan merupakan cara untuk mengorganisasikan dokumen penerbangan angkutan udara. Dokumen penerbangan yang memiliki isi yang sama akan dikelompokkan ke dalam kategori yang sama. Dengan demikian orang yang melakukan pencarian informasi yang berhubungan dengan data angkutan udara dapat dengan mudah melewati kategori yang tidak berkaitan dengan informasi penerbangan. Tahapan pengklasifikasian data penerbangan yaitu menggunakan model dari training data penerbangan angkutan udara yang menghasilkan prediksi

## 5. KESIMPULAN

1. Dari data angkutan udara penerbangan domestik dan penerbangan internasional di bandara Sultan Iskandar Muda selama periode bulan januari 2010 hingga bulan maret 2011 terdapat 12 variabel yang mempengaruhi data penerbangan domestik dan penerbangan internasional dengan 22 training data (75%) dengan metode support vector machine (SVM) yang memiliki 4 support vector pada garis hyperplane terbaik yaitu 2 model untuk penerbangan domestik dan 2 model untuk penerbangan internasional.
2. Pada proses prediksi data model dilakukan dengan 8 testing data (25%) berdasarkan kurva ROC oleh kernel linear dengan tingkat akurasi model hingga 84,31%. Artinya tingkat keakuratan model mendekati 100% dan dapat dipercaya.
  - Arus penerbangan domestik lebih mendominasi pada kunjungan wisatawan ke banda aceh bila dibandingkan dengan arus kunjungan wisatawan mancanegara (wisatawan internasional) sehingga diharapkan Pemerintah kota banda aceh beserta

instansi terkait lainnya khususnya bandara sultan iskandar muda dapat saling bekerjasama lebih giat dan bekerja lebih keras lagi dalam mempromosikan potensi wisata di aceh sehingga diharapkan kunjungan wisatawan mancanegara lebih banyak lagi di tahun mendatang dan dapat menambah pemasukan devisa negara dan pendapatan masyarakat dari sektor wisata seiring dengan program tahun kunjungan wisata di tahun 2011 dan program tahun kunjungan wisata dapat terus berkesinambungan dan dilanjutkan pada tahun berikutnya.

## REFERENSI

- Beeza-Yates, Ricada & Berthier Ribeiro-Neto. 1999.,** *Modern Information Retrieval.* Addison Wesley.
- Brefeld, Ulf., 2005,** *AUC Maximizing Support Vector Learning,*
- Chin, 1998,** *Using a Radial Basis Function as Kernel,*  
<http://svrwww.eng.cam.ac.uk/~kkc21/thesi-main/node31.html>.K.K
- Christianini, N. And Shawe Taylor, J., 2000,** *An Introduction to Support Vector Machine and other Kernel Based Learning methods,* Cambridge University Press.
- Cover Knowledge Base\_2006.** *Understanding Stemiry,*  
<http://www.Cover.com/and-support/articles/information-CES4-060330-3-Understanding,Steming>.
- Dimitriadou, Evgenia, Hornik, Kurt, Leisch, Friedrich, Meyer, David, and Andreas, Weingessel, 2007,** *E1071 : Misc Functions of the Department of Statistics (e1071),* TU Wien. R package version 1.5-17.
- Feldman, Susan. 2004,** *Why Categorize*  
<http://www.kmworld.com/Articles/Editorial/Feature/why-categorize-3f-9580.aspx>
- Goller, C et. Al. 2000.** *Automatic Document Classification : A Thorough Evaluation of Various Methods.*

- Han, J. and Kamber, M. 2006.** *Data Mining Concepts and Techniques*, Second Edition. Morgan Kauffman. San Francisco.
- Joachims, Thorstem., 1999.** *Transductive Inference for Text Classification Using Support Vector Machines. Proceedings of the International and Conference on Machine learning (ICML)*
- Karatzoglou, Alexandros., 2006,** *Support Vector Machines in R*, <http://www.jstatsoft.org/>
- MedCalc Software bvba. .ROC Curve Analysis : Introduction.** MedCalc Software, Broekstraat 52, 9030 Mariakerke. Belgium. <http://www.medcalc.be/manual/roc.php>.
- Nugroho, Anto Satrio., 2003,** *Support Vector Machine Teori dan Aplikasinya dalam Bioinformatika.*
- Pramudiono, I. 2007,** *Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data.* <http://www.ilmucomputer.org/wpcontent/uploads/2006/08/IK0datamining.zip>.
- Rainardi, Vincent. 2008.** *Building a Data Warehouse with Examples in SQL Server.* Springer. New York.
- R.Goldstein, Darlene. 2003.** *ROC Curves, Swiss Institute of Bioinformatics, Switzerland*
- Santosa, Budi. 2007.** *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis.* Graha Ilmu. Yogyakarta.
- Sebastiani, Fabrizio.2002.** *Machine Learning in Automated Text Categorization.* ACM Computing Surveys, 34 (1) : 1-47 <http://nmis.isti.cnr.it/sebastiani/publications/ACMCS02.pdf>.
- Sembiring, Krisantus. 2007.** *Tutorial SVM Bahasa Indonesia.* Skripsi, S1 Teknik Informatika, Teknik Elektro dan Informatika, ITB. Bandung
- Sun, Aixin, 2002.** *Web Classification Using Support Vector Machine.* Proceedings of the 4<sup>th</sup> Int. Workshop on Web Information and Data Management (WIDM 2002) held in conj. With CIKM 2002, Virginia. USA.
- Tuszynski, Jarek. 2007.** *caTools : Tools : moving window statistics, GIF, Base64, ROCAUC, etc.. R package version 1.8.*
- Vapnik, V, 1995.** *The Nature of Statistical Learning Theory*, Springer-Verlag.

**Witten, I. H and Frank, E. 2005.** *Data Mining : Practical Machine Learning Tools and Techniques.* Second Edition. Morgan Kauffman. San Francisco