

ANALISIS DAN IMPLEMENTASI PERBANDINGAN ALGORITMA KNN (K-NEAREST NEIGHBOR) DENGAN SVM (SUPPORT VECTOR MACHINE) UNTUK PREDIKSI PENAWARAN PRODUK

Comparative Analysis and Implementation of KNN (K-Nearest Neighbor) with SVM (Support Vector Machine) Algorithm for Product Offering Prediction

Rizki Djuwi Probo¹, Budhi Irawan², R. Rumani M.³

^{1,2,3}Prodi S1 Sistem Komputer, Fakultas Teknik Elektro, Universitas Telkom

¹rizkidjuwiprobo@gmail.com, ²bir@telkomuniversity.ac.id, ³rumani@telkomuniversity.ac.id

Abstrak

Di dalam perusahaan salah satu hal yang paling penting adalah produk mereka laku dengan keuntungan yang sebesar-besarnya. Salah satu contoh cara memasarkannya yaitu dengan teknik *Telemarketing*. *Telemarketing* digunakan oleh perusahaan karena efisien dari sisi waktu yang digunakan dan uang yang dikeluarkan lebih kecil dibandingkan berjualan secara langsung. Akan tetapi menjadi sia-sia bila perusahaan tidak dapat memprediksi penjualan produknya. Saat ini, *history* jumlah penawaran produk *telemarketing* yang dimiliki sudah banyak, sulit jika menggunakan teknik konvensional dalam menghitung dan melakukan prediksi penawaran produk.

Dalam penelitian ini, untuk mengatasi masalah tersebut maka diuji menggunakan 2 buah metode algoritma yaitu Algoritma KNN (*K-Nearest Neighbors*) dan SVM (*Support Vector Machine*). Dari kedua buah metode tersebut akan dikaji metode mana yang paling akurat digunakan untuk memprediksi penawaran produk sehingga mengoptimalkan penjualannya. Dengan adanya prediksi penawaran produk tersebut diharapkan dapat mempermudah dan menambah nilai dalam melakukan optimasi pada bagian *telemarketing* untuk mendukung strategi penjualan berdasarkan data yang dihasilkan dari proses yang ada.

Hasil penelitian ini memudahkan bagian *telemarketing* dalam melaksanakan tugasnya. Nilai akurasi yang didapat dari aplikasi ini membantu memprediksi nasabah yang akan menggunakan produk banknya. Dari pengujian yang dilakukan dalam studi kasus ini, algoritma SVM lebih unggul dibandingkan dengan algoritma KNN. Nilai akurasi algoritma SVM pada 266 data sebesar 69,55%, sedangkan nilai akurasi KNN sebesar 49,62%.

Kata kunci: Data Mining, KNN, SVM

ABSTRACT

In the company one of the most important things is to gain profit as much as possible from their products. One of the example to market it is by telemarketing technique. Telemarketing is used by the company because it is very efficient in terms of time spent and money spent is smaller than sell directly. But became vain if the company is unable to predict product sales. Today, the number of product offering history is too many, it is difficult when using conventional techniques in calculating and predicting product offerings.

In this study, to resolve the issue it will be tested by using 2 algorithm methods which are KNN (K-Nearest Neighbor) algorithm and SVM (support Vector Machine) algorithm. From this two methods will be studied which method has most accurately used to predict product offerings that optimize the seller. With the prediction of product offerings are expected to simplify and add value in optimizing the telemarketing department to support sales strategy based on data generated from the existing process.

The results of this study will facilitate telemarketing department in carrying out their duties. Because the value of accuracy obtained from these applications help predict which customers will use the product of the bank. From the tests carried out in this case study, SVM algorithm is superior compared with KNN algorithm. SVM algorithm accuracy value at 266 data at 69.55%, while the value of KNN accuracy of 49.62%.

Keywords : Data Mining, KNN, SVM

1. Pendahuluan

PT. BANK XYZ Tbk merupakan salah satu bank terkenal di Indonesia yang menyediakan produk dan layanan Perbankan. Salah satu unit kerja dari PT. BANK XYZ adalah bagian *Telemarketing* yang mempunyai tugas melakukan penjualan melalui telepon terhadap *database* prospek. *Telemarketing* sering digunakan oleh banyak perusahaan karena efisien dari sisi waktu yang digunakan dan uang yang dikeluarkan lebih sedikit dibandingkan berjualan secara langsung. Akan tetapi menjadi sia-sia bila perusahaan tidak dapat memprediksi penjualan produknya. Sehingga dengan adanya prediksi tersebut perusahaan dapat membuat strategi penjualan yang akan dibuat seperti penawaran produk dilakukan oleh *agent* siapa, jam telepon yang baik bagi pegawai swasta dan lain - lain.

Saat ini, *history* jumlah penawaran produk *telemarketing* yang dimiliki sudah banyak, sulit jika menggunakan teknik konvensional dalam menghitung dan melakukan prediksi penawaran produk. Untuk mengatasi masalah tersebut, maka diuji menggunakan 2 buah metode algoritma yaitu Algoritma KNN dan SVM yang akan dikaji metode mana yang paling akurat digunakan untuk memprediksi penawaran produk pada PT Bank XYZ Tbk sehingga mengoptimalkan penjualannya. Setelah itu semua dibuatkan *prototype* prediksi penawaran produk yang diharapkan dapat mempermudah dan menambah nilai dari PT Bank XYZ Tbk dalam melakukan optimasi pada bagian *telemarketing* untuk mendukung strategi penjualan berdasarkan data yang dihasilkan dari proses yang ada.

2. Tinjauan Pustaka

2.1 Data Mining

Secara sederhana *data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar ^[1]. *Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data ^[10]. *Data mining*, sering juga disebut sebagai *knowledge discovery in database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar ^[12].

2.2 Klasifikasi

Secara sederhana *data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar ^[1]. *Data mining* juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data ^[10]. *Data mining*, sering juga disebut sebagai *knowledge discovery in database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar ^[12].

2.3 Algoritma KNN (K-Nearest Neighbor)

Algoritma *k-Nearest Neighbour* (KNN) adalah algoritma pengklasifikasian data sederhana dimana penghitungan jarak terpendek dijadikan ukuran untuk mengklasifikasikan suatu kasus baru berdasarkan ukuran kemiripan. Algoritma KNN tergolong dalam algoritma *supervised* yaitu proses pembentukan algoritma diperoleh melalui proses pembelajaran (*learning*) pada *record-record* lama yang sudah terklasifikasi dan hasil pembelajaran tersebut dipakai untuk mengklasifikasikan *record* baru dengan *output* yang belum diketahui ^[9].

Rumus *Euclidian* :

$$d_i = \sqrt{\sum_{j=1}^p (x_{ij} - x_{kj})^2}$$

Keterangan:

x_1 = Sampel Data
 x_2 = Data Uji / Testing
 i = Variabel Data
 d = Jarak
 p = Dimensi Data

2.4 Algoritma SVM (Support Vector Machine)

Algoritma *Support Vector Machine* (SVM) adalah algoritma klasifikasi yang sering digunakan sejak ditemukan pada tahun 1995. Proses klasifikasi sering kali melibatkan pemisahan data menjadi data latih dan data uji. Setiap *instances* pada set data latih memiliki satu *target value* dan beberapa atribut. Tujuan dari SVM adalah untuk menghasilkan model berdasarkan data latih, yang dapat memberikan prediksi *target value* dari data uji yang diberikan hanya dengan melihat atributnya^[5].

2.5 Confusion Matrix

Confusion Matrix adalah alat yang berguna untuk menganalisis seberapa baik *classifier* mengenali *tuple* dari kelas yang berbeda. TP dan TN memberikan informasi ketika *classifier* benar, sedangkan FP dan FN memberitahu ketika *classifier* salah. Contoh gambar *Confusion matrix*^[3].

		Predicted class		
		Yes	No	Total
Actual class	Yes	TP	FN	P
	No	FP	TN	N
	Total	P'	N'	P+N

3. Analisa dan Perancangan Sistem

3.1 Deskripsi Umum Sistem

Jika implementasi dari perancangan berhasil dilakukan, sistem akan memiliki spesifikasi sebagai berikut ini:

1. Mampu melakukan proses "*Learning*", dimana sistem mempelajari alur algoritma KNN dan SVM dalam mengklasifikasikan data;
2. Mampu melakukan proses "*Forecasting*", dimana sistem mampu melakukan prediksi dari data yang disajikan berdasarkan proses "*Learning*" sebelumnya;
3. Mampu membandingkan performa kedua algoritma melalui nilai akurasi.

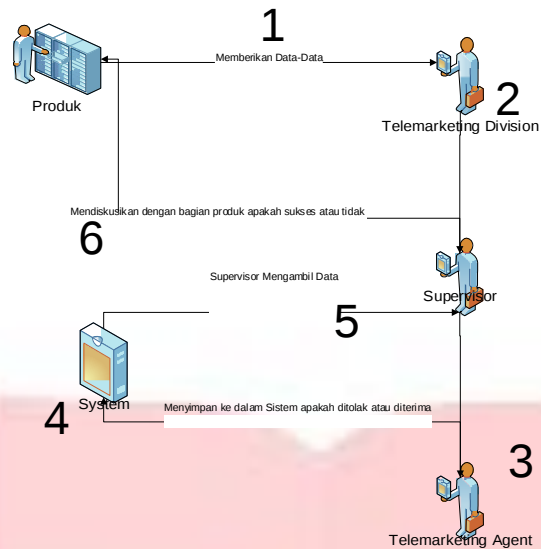
Implementasi dilakukan dengan membuat sebuah aplikasi yang dapat digunakan pengguna untuk memprediksi penawaran produk berdasarkan data penawaran produk periode November 2014. Masukan aplikasi berupa file csv yang berisi 667 data penawaran produk yang telah dilakukan oleh divisi *Telemarketing* sebelumnya. Keluaran yang diharapkan dari sistem berupa daftar prediksi penawaran produk yang telah diolah oleh algoritma dengan nilai akurasi terbaik.

3.2 Perancangan Sistem Umum

Perancangan dimulai dengan kegiatan pengumpulan data berupa 1000 data *history* penawaran produk selama bulan November 2014. Setelah seluruh data didapatkan, maka dilakukan perancangan model *data mining* dengan teknik CRISP-DM. Model CRISP-DM (*Cross – Industry Standard Process for Data Mining*) yang terdiri dari 6 tahap proses, yaitu:

1. *Business Understanding*
2. *Data Understanding*
3. *Data preparation*
4. *Modeling*
5. *Evaluation*

6. Deployment

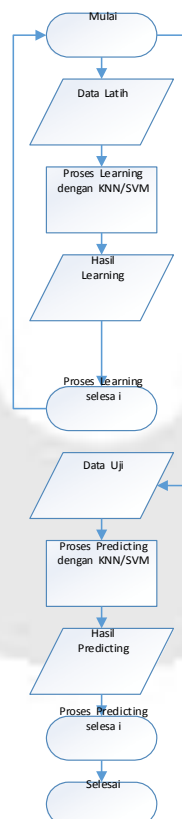


Gambar 3.1 Sistem yang sedang berjalan saat ini

3.3

Diagram Alir Sistem

Gambaran umum dari perancangan aplikasi dapat dilihat pada diagram alir berikut:



Gambar 3.2 Diagram Alir Sistem

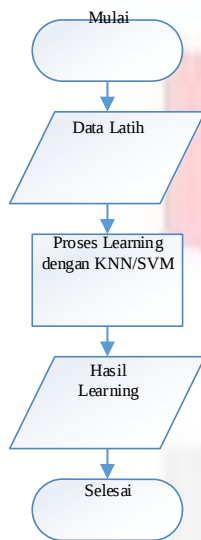
Dari diagram alir di atas dapat diketahui bahwa sistem dimulai dengan masukan data latih yang akan diolah dengan algoritma KNN dan SVM pada proses “*Learning*”. Data latih merupakan data *History* Penawaran Produk Kredit Tanpa Agunan periode November 2014 yang diambil sebanyak 667 data. Data latih tersebut diolah dengan algoritma KNN dan SVM pada proses *learning* untuk menghasilkan Hasil *Learning* yang nantinya akan menjadi acuan untuk proses *predicting*. Jika Hasil *Learning* muncul pada aplikasi, berarti proses *learning* telah selesai.

Kemudian, pengguna dapat melanjutkan ke proses *predicting* yang dimulai dengan memasukkan input berupa Data Uji. Data Uji diambil dari kumpulan data History Penawaran Produk Kredit Tanpa Agunan periode Noveber 2014 sebanyak 333 data. Setelah data uji diolah, keluaran yang dihasilkan adalah daftar prediksi penawaran produk berupa data Hasil *Predicting*. Ketika aplikasi telah menunjukkan Hasil *Predicting*, berarti proses *predicting* telah selesai.

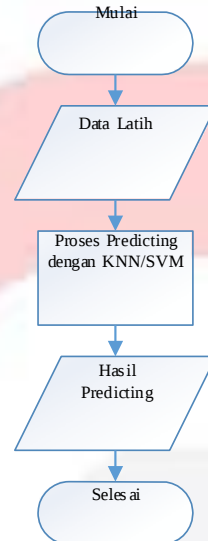
Hasil *Learning* dan Hasil *Predicting* kemudian akan menjadi masukan untuk proses penghitungan nilai akurasi dengan *Confusion Matrix*. Keluaran dari proses perhitungan nilai akurasi tersebut adalah persentase nilai akurasi dari masing-masing algoritma.

3.4 Proses Learning

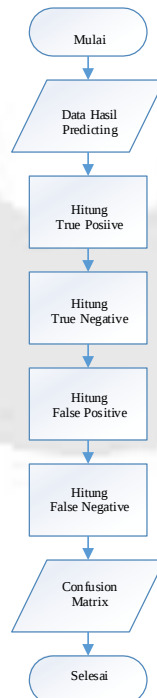
Proses *learning* pada aplikasi dapat dilihat pada diagram alir berikut:



Gambar 3.3 Diagram Proses Learning



Gambar 3.4 Diagram alir proses *predicting*



Gambar 3.5 Diagram alir proses *Confusion Matrix*

4. Implementasi dan Pengujian Sistem

4.1 Implementasi Antarmuka

Berdasarkan kebutuhan masukan dan keluaran pada gambaran umum aplikasi, perancangan antarmuka dari aplikasi dapat digambarkan sebagai berikut ini:

Gambar 4.1 Perancangan antarmuka Prediksi Penawaran Produk

4.2 Pengujian dengan 100 data

Tabel 4.1 Tabel *Confusion Matrix* untuk 100 data

KNN	YES	NO
YES	23	22
NO	21	34

Sensitifitas 51.11%
Spesifisitas 61.82%

Akurasi 57.00%
Error Rate 43.00%

SVM	YES	NO
YES	53	1
NO	36	10

Sensitifitas 98,15%
Spesifisitas 21,74%

Akurasi 63,00%
Error Rate 37,00%

4.3 Pengujian dengan 266 data

Tabel 4.2 Tabel *Confusion Matrix* untuk 266 data

KNN	YES	NO
YES	69	54
NO	80	63

Sensitifitas 56.10%
Spesifisitas 44.06%

Akurasi 49.62%
Error Rate 50.38%

SVM	YES	NO
YES	130	19
NO	62	55

Sensitifitas 87,25%
Spesifisitas 47,01%

Akurasi 69,55%
Error Rate 30,45%

5. Kesimpulan dan Saran

5.1 Kesimpulan

Berdasarkan pengujian yang dilakukan maka didapatkan kesimpulan sebagai berikut:

1. Tingkat akurasi yang ditunjukkan oleh algoritma SVM lebih baik dibandingkan dengan algoritma KNN, nilai akurasi algoritma SVM pada kasus ini di pengujian 100 dan 266 data sebesar 63% dan 69,55% sedangkan nilai akurasi algoritma KNN pada kasus ini di pengujian 100 dan 266 data sebesar 57% dan 49,62%
2. Tingkat kesalahan yang ditunjukkan oleh algoritma SVM lebih kecil dibandingkan dengan algoritma KNN, tingkat kesalahan algoritma SVM pada kasus ini di pengujian 100 dan 266 data sebesar 37% dan 30,45% sedangkan tingkat akurasi algoritma KNN pada kasus ini di pengujian 100 dan 266 data lebih besar yaitu 43% dan 50,38%.
3. Jika dilihat dari banyaknya algoritma yang melakukan kesalahan klasifikasi, maka algoritma SVM lebih baik dibandingkan dengan algoritma KNN pada kasus ini, karena algoritma KNN melakukan kesalahan klasifikasi pada pengujian 100 dan 266 data sebanyak 43 dan 134 data, sedangkan algoritma SVM melakukan kesalahan klasifikasi pada pengujian 100 dan 266 data sebanyak 37 dan 81 data.

5.2 Saran

Dari aplikasi yang dibangun masih terdapat banyak kekurangan dan perlu pengembangan untuk mendapatkan aplikasi yang lebih baik. Adapun saran untuk pengembangan aplikasi ini sebagai berikut:

1. Pengembang aplikasi berikutnya disarankan untuk menggunakan algoritma yang belum dibahas pada aplikasi ini.
2. Pengembang disarankan mencari alternatif cara optimasi penggunaan algoritma data mining agar tingkat prediksi lebih optimal.

DAFTAR PUSTAKA

- [1] Davies., Beynon, Paul., "Database Systems Third Edition", Palgrave Macmillan, New York, 2004.
- [2] Gorunescu, Florin. 2011. *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg : Springer
- [3] Han, J. and Kamber, M, 2006, "Data Mining Concepts and Techniques Second Edition". Morgan Kauffman, San Francisco.
- [4] Hidayatulloh, Taufik. "Kajian Komparasi Penerapan Algoritma Support Vector Machine (SVM) Dan Multilayer Perceptron (MLP) Dalam Prediksi Indeks Saham Sektor Perbankan: Studi Kasus Saham LQ45 IDX Bank BCA." Prosiding SNIT (2014).
- [5] Hsu, C. W., Chang, C. C., & Lin, C. J., *A practical guide to support vector classification*, 2003.
- [6] Huda, Nuqson Masykur. *Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa (Studi Kasus di Fakultas MIPA Universitas Diponegoro)*. Diss. FACULTY OF MATHEMATICS AND NATURAL SCIENCES, 2010.
- [7] Kustiyahningsih, Yeni, and Nikmatus Syafa'ah. *Sistem Pendukung Keputusan Untuk Menentukan Jurusan Pada Siswa SMA Menggunakan Metode KNN dan SMART*. *Jurnal Sistem Informasi Indonesia* 1.1 (2015).
- [8] Larose, Daniel T. *Data mining methods & models*. John Wiley & Sons, 2006.
- [9] Leidiyana, Henny. "Penerapan Algoritma K-Nearest Neighbor untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor." PIKSEL (Penelitian Ilmu Komputer Sistem Embedded dan Logic) 1.1 (2013).
- [10] LUKMAN, LUKMAN. "PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) DALAM PEMILIHAN BEASISWA: STUDI KASUS SMK YAPIMDA." *Faktor Exacta* 9.1 (2016): 49-57.

- [11] Pramudiono, I. 2007. Pengantar Data Mining : Menambang Permata Pengetahuan di Gunung Data. <http://www.ilmukomputer.org/wp-content/uploads/2006/08/iko-datamining.zip> Diakses pada tanggal 20 Juli 2016 jam 08.54
- [12] Rismawan, T. I., Prabowo, A.W., Kusumadewi, S., *Sistem Pendukung Keputusan Berbasis Pocket PC sebagai Penentu Status Gizi Menggunakan Metode KNN (K –Nearest Neighbor)*. *Jurnal Teknoin I*, No.2, Vol.12 (2008).
- [13] Santosa, Budi. 2007. “Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis”. Graha Ilmu, Yogyakarta.
- [14] Saputro, Nugroho Dwi. *Penerapan Algoritma Support Vector Machine untuk Prediksi Harga Emas*. *JIU* 1.1 Juni (2015).
- [15] SYAMSIAH, SYAMSIAH. "PEMILIHAN MODEL PENENTUAN KELAYAKAN PINJAMAN ANGGOTA KOPERASI BERDASARKAN ALGORITMA SUPPORT VECTOR MACHINE, GENETIC ALGORITHMS, DAN NEURAL NETWORK." *Faktor Exacta* 7.2 (2015): 141-153.
- [16] Whidha, Ayu Primastuti, Yati Rohayati, Mochamad Rian Murtiadi. *Pemilihan Sasaran Cross Selling Layanan Internet Speedy Pada Pelanggan Telepon Rumah Menggunakan Algoritma k-Nearest Neighbor*. Seminar Nasional Aplikasi Teknologi Informasi (SNATI), Bandung: Institut Teknologi Telkom. (2012).
- [17] Witten, I. H and Frank, E. 2005. *Data Mining : Practical Machine Learning Tools and Techniques* Second Edition. Morgan Kauffman : San Francisco.
- [18] Yofianto, E., 2010,. K-Nearest Neighbor, <http://kuliahinformatika.wordpress.com/2010/02/13/buku-ta-k-nearest-neighbor-knn/.htm> , diakses 3 Desember 2015.