ANALISIS PENERAPAN ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI PRIORITAS PENGEMBANGAN JALAN DI PROVINSI KALIMANTAN SELATAN

Erfan Karyadiputra[1)], Nadiya Hijriana[2)]

[1]Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari (Erfan Karyadiputra)
email : Erfantsy@gmail.com
[1]Fakultas Teknologi Informasi, Universitas Islam Kalimantan Muhammad Arsyad Al Banjari (Nadiya Hijriana)
email : Nadyahijriana@gmail.com

Abstrak

Provinsi Kalimantan Selatan secara administrasi wilayah terdiri atas 2 kota dan 11 Kabupaten. Pertumbuhan perekonomian di Provinsi Kalimantan Selatan sangat dipengaruhi oleh lancarnya distribusi barang dan jasa. Oleh sebab itu sektor perhubungan terutama infrastruktur jalan memegang peranan sangat penting sebagai prasarana kelancaran suatu distribusi barang dan jasa. Kondisi infrastruktur ruas jalan yang rusak dapat menghambat pendistribusian barang dan jasa sehingga perlu dilakukan perbaikan-perbaikan terhadap jalan tersebut agar distribusi barang dan jasa menjadi lancar. Namun, usaha tersebut terkendala dengan anggaran yang terbatas sehingga instasi terkait harus menentukan daftar prioritas jalan yang akan diperbaiki. Salah satu teknik yang dapat diterapkan yaitu dengan melakukan analisa dengan memanfaatkan teknik data mining berupa penerapan algoritma Naive Bayes Classification untuk mengklasifikasi daftar prioritas pengembangan jalan di Provinsi Kalimantan Selatan. Penerapan algortima naive bayes pada dataset menghasilkan akurasi sebesar 88.27% dan AUC sebesar 0.933.

***Keywords** : Data Mining, Klasifikasi, Naive Bayes*

1. PENDAHULUAN

Provinsi Kalimantan Selatan terbentuk pada tahun 1956 berdasarkan Undang-Undang No.5 yang secara administrasi wilayahnya terdiri atas 2 kota yaitu Banjarmasin dan Banjarbaru serta 11 kabupaten yaitu Barito Kuala, Tanah Laut, Kotabaru, Banjar, Tapin, Hulu Sungai Selatan, Hulu Sungai Tengah, Hulu Sungai Utara, Tabalong, Tanah Bumbu dan Balangan. Provinsi Kalimantan Selatan memiliki peran penting dalam perekonomian nasional terutama dalam sektor pertanian, perkebunan, peternakan, kehutanan dan perikanan. Provinsi Kalimantan Selatan juga memiliki komuditas utama yaitu kelapa sawit dan karet. Pertumbuhan perekonomian di Provinsi Kalimantan Selatan sangat dipengaruhi oleh lancarnya distribusi barang dan jasa. Oleh sebab itu sektor perhubungan terutama infrastruktur jalan memegang peranan sangat penting sebagai prasarana kelancaran suatu distribusi barang dan jasa. Kondisi infrastruktur ruas jalan yang rusak dapat menghambat pendistribusian barang dan jasa sehingga perlu dilakukan perbaikan-perbaikan terhadap jalan tersebut agar distribusi barang dan jasa menjadi lancar. Namun, usaha tersebut terkendala dengan anggaran yang terbatas sehingga instasi terkait harus menentukan daftar prioritas jalan yang akan diperbaiki. Oleh sebab itu, diperlukan analisa dengan memanfaatkan teknik data mining berupa penerapan algoritma *Naive Bayes Classification* untuk mengklasifikasi daftar prioritas pengembangan jalan di Provinsi Kalimantan Selatan.

2. METODE PENELITIAN

a. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah berasal dari hasil pendataan jalan dan jembatan di Provinsi kalimantan Selatan. Variabel yang akan digunakan pada penelitian ini adalah :

Tabel 1. Variabel Data Jalan Provinsi

Variabel	Keterangan	Kategori
Y	Status	Disetujui, Ditolak
X1	Klasifikasi Jalan	Jalan Provinsi, Jalan Nasional
X2	Jenis Pengerasan	Aspal, Agregat, Tanah
X3	Kelas Jalan	K1, K2
X4	Kecepatan	-
X5	Jumlah Lobang	-
X6	Titik Longsor	-
X7	Jumlah Longsor (m)	-
X8	Rusak (Km)	-
X9	Sedang (Km)	-
X10	Baik (Km)	-
X11	Panjang (Km)	-
X12	Lebar (m)	-
X13	Tipe Jalan	2/2D, 2/2 UD, 4/2 D, 4/2 UD, 2/1 D, 2/1 UD
X14	Alinyemen Horizontal	Sedikit Tikungan, Lurus, Banyak Tikungan
X15	Alinyemen Vertikal	Bukit, Datar, Pegunungan
X16	Lalu Lintas	Sangat Tinggi, Tinggi, Sedang, Rendah, Sangat Rendah
X17	Tata Guna Lahan	Pemukiman, Perkotaan, Areal Persawahan, Ruko, Lahan Kosong, Pertanian, Perkantoran, Daerah Wisata, Perkebunan, Tempat Wisata, Pertambangan
X18	Kondisi	Baik, Sedang, Rusak, Rusak Ringan, Rusak Berat

b. Pengolahan Data Awal

Tahapan selanjutnya setelah pengumpulan data maka data tersebut kemudian diolah dengan melakukan *pre-processing* seperti identifikasi dan inspeksi sejumlah atribut-atribut yang tidak relevan. Kemudian setelah diketahui sejumlah atribut yang tidak relevan maka tahap selanjutnya adalah mengeliminasi sejumlah atribut tersebut dan melakukan subtitusi data sehingga dapat diproses dalam *data mining*.

c. Eksperimen dan Pengujian Metode

Terdapat beberapa tahapan dalam melakukan eksperimen, yaitu :

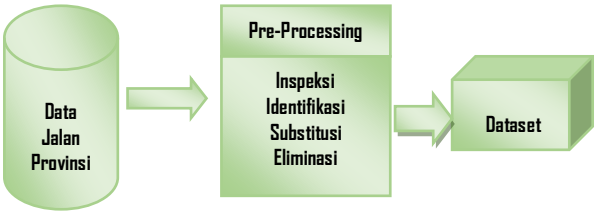
Tahap 1 : Tahap pertama pengujian klasifikasi pertama algoritma *naive bayes* menggunakan dataset dengan melakukan pergantian beberapa jenis *parameter* pada *X-validation* yaitu *sampling type* yang terdiri atas tiga jenis, diantaranya *linear sampling*, *shuffled sampling*, dan *stratified sampling* sehingga didapatkan *sampling type* yang sesuai untuk digunakan pada data yang diuji

Tahap 2 : Tahap selanjutnya memisahkan dataset menjadi 2 jenis data, yakni data *training* dan data *testing* dengan menggunakan pembagian persentase dari jumlah dataset. Apabila data *training* yang digunakan sebanyak 10% dari dataset maka persentase data *testing* adalah 90%. Dan apabila data *training* yang digunakan sebanyak 20% dari dataset maka persentase data *testing* adalah 80%. Jumlah data *training* apabila dijumlah dengan data *testing* akan menjadi jumlah dataset tersebut. Jumlah persentase pada data *trainning* nantinya akan ditentukan berdasarkan perulangan pada proses *cross-validation*. Contohnya, proses *cross-validation* pada iterasi ke 4 dari total 10 iterasi, maka persentase jumlah data *training* adalah 40% dan persentase jumlah data *testing* adalah 60%. Proses pemisahan data *training* dan data *testing* mengikuti proses *X-validation* dengan jumlah iterasi mulai dari 2,3,4,5,6,7,8,9 dan 10 dan agar data yang digunakan pada proses itu tetap maka akan menggunakan dataset yang sama.

Tahap 3 : Mengevaluasi dan melakukan pengujian *performance* dengan menggunakan *confusion matrix* dan *Compare* ROC sehingga dapat diketahui perbandingan hasil akurasi dan nilai AUC. Nilai AUC digunakan untuk menentukan hasil klasifikasi kedalam klasifikasi sangat baik, klasifikasi baik, klasifikasi cukup, klasifikasi buruk dan klasifikasi salah.

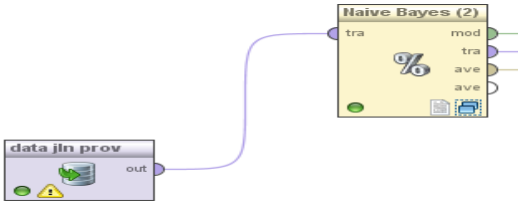
3. HASIL DAN PEMBAHASAN

Sebelum melakukan eksperimen dan pengujian model, data-data yang dikumpulkan terlebih dahulu diolah agar dapat diproses dalam *data mining*.



Gambar 1. Pre-Processing

Tahap pertama pengujian klasifikasi pertama algoritma *Naive Bayes* menggunakan dataset dengan melakukan pergantian beberapa jenis *parameter* pada *X-Validation* yaitu *Sampling Type* yang terdiri atas tiga jenis, diantaranya *Linear Sampling, Shuffled Sampling*, dan *Stratified Sampling* sehingga didapatkan *Sampling Type* yang sesuai untuk digunakan pada data yang diuji



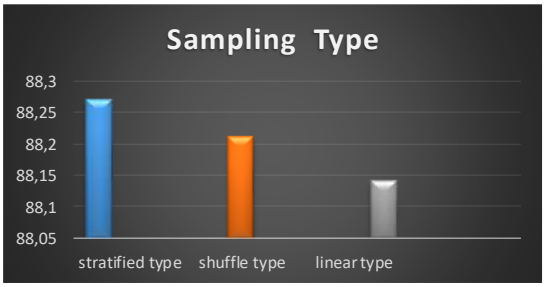
Gambar 2. Pengujian Tipe Sampling

Ketika diimplementasi menghasilkan data sebagai berikut :

Tabel 2. Hasil Pengujian Tipe Sampling

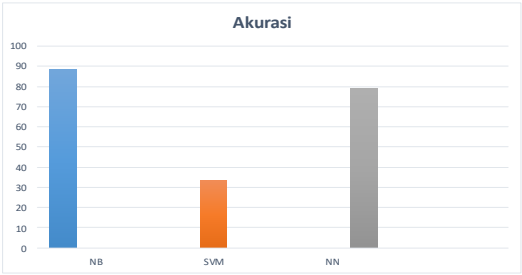
Sampling Type	***Linear Sampling***	***Shuffled Sampling***	***Stratified Sampling***
Accuracy	83.18	85.41	85.80

Dari pengujian tersebut menghasilkan penggunaan *X-validation* dengan *sampling type stratified* memiliki tingkat akurasi yang sedikit lebih baik daripada *sampling type* yang lain. Perbandingan akurasi dari beberapa tipe validasi dalam bentuk grafik seperti gambar dibawah ini



Gambar 2. Pengujian Tipe Sampling

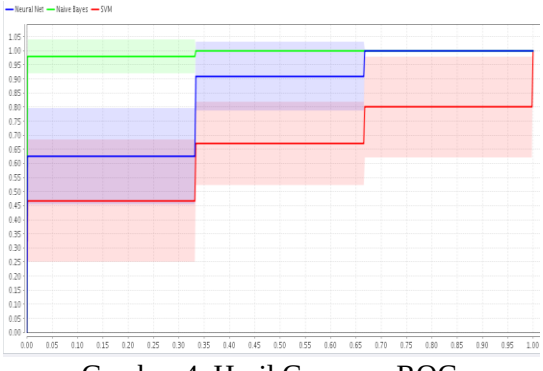
Berdasarkan penelitian dari berbagai percobaan yang ekstensif dan pembuktian teoritis, menunjukkan bahwa penggunaan 10-*fold cross-validation* adalah pilihan terbaik untuk mendapatkan hasil validasi yang akurat sehingga pada pengujian tersebut didapatkan akurasi 88,27% dengan nilai AUC sebesar 0,933. Kemudian untuk mendapatkan metode terbaik maka hasil *performance Naive Bayes* tersebut dibandingkan dengan algoritma lainnya *seperti Support Vector Machine* (SVM) dan *Neural Networks* menggunakan *Confusion Matrix* dan *Compare* ROC sehingga dapat terlihat perbandingan nilai akurasi dan didapatkan model terbaik terkhusus dalam melakukan klasifikasi daftar prioritas pengembangan jalan provinsi selatan



Gambar 3. Hasil Perbandingan Akurasi

Berdasarkan hasil diatas terlihat *Naive Bayes* lebih baik berdasarkan keunggulan tingkat akurasi sebesar 88,27% jika dibandingkan dengan algoritma *Support Vector Machine* (33.65%) dan *Neural Networks* (78.97%).

Pengujian selanjutnya menggunakan *Compare* ROC. Hasil perbandingan tingkat kehandalan klasifikasi seperti gambar dibawah ini



Gambar 4. Hasil Compare ROC

Dari pengujian tersebut dapat dianalisa bahwa *Naive Bayes* lebih baik dan termasuk *excellent classification* berdasarkan keunggulan dalam hal kehandalan dalam klasifikasi dengan nilai sebesar 0,933 karena memiliki nilai AUC yang lebih baik daripada algoritma *Support Vector Machine dan Neural Networks.*

4. KESIMPULAN

Dari hasil penelitian yang dilakukan dari tahap awal pengujian hingga akhir pengujian, dan dari hasil perbandingan dapat disimpulkan bahwa secara umum *Naive Bayes* lebih baik jika dibandingkan dengan algoritma *Support Vector Machine* dan *Neural Networks* berdasarkan keunggulan tingkat akurasi sebagai tingkat kedekatan antara nilai klasifikasi dengan nilai aktual dan berdasarkan keunggulan dalam hal kehandalan dalam klasifikasi karena memiliki nilai AUC yang lebih baik sehingga *Naive Bayes* dapat diterapkan dalam mengklasifikasi daftar prioritas pengembangan jalan Provinsi Kalimantan Selatan dengan akurasi 88,27% dan termasuk kedalam *excellent classification* dengan nilai AUC sebesar 0.930.

5. REFERENSI

[1] V. Moertini, Data Mining Sebagai Solusi Bisnis, Integral Vol. 7 No. 1, 2002.

[2] J. Han dan M. Kamber, Data Mining : Concepts and Techniques, Second penyunt., M. R. Jim Gray, Penyunt., San Francisco: Morgan Kaufmann Publishers, 2007.

[3] W. Ian H. dan E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2 penyunt., San Francisco: Morgan Kaufmann Publishers is an imprint of Elsevier, 2005.

[4] F. Gorunescu, Data Mining: Concept, Models and Techniques, Romania: Springer, 2010.

[5] Larose, Data Mining Methods And Models, Canada: John Wiley & Sons, Inc, 2006.

[6] R. IT, "data mining data preprocessing," 8 3 2010. [Online]. Available: https://rencanait.wordpress.com. [Diakses 1 2 2015].

[7] D. Aprilla, D. A. Baskoro, L. Ambarwati dan I. W. S. Wicaksana, "Belajar Data Mining dengan RapidMiner," Jakarta, academia.edu, 2013.

[8] C. Sammut dan G. Web, Encyclopedia of machine learning, New York: Springer, 2011.

[9] Zhang dan Wang, Application of Bayesian Method to Spam SMS Filtering, International Conference on Information Engineering and Computer Science, 2009.

[10] J. E. Simatupang, Studi Penentuan Prioritas Peningkatan Ruas Jalan Nasional Di Provinsi Kalimantan Selatan, Malang: Brawijaya, 2011.

[11] A. Suwandi, Evaluasi Tingkat Kerusakan Jalan Dengan Metode Pavement Condition Index (PCI) Untuk Menunjang Pengambilan Keputusan, Yogyakarta: UGM, 2011.

[12] Undang-Undang, Undang-Undang Republik Indonesia No 22 Tahun 2009, Jakarta: Dirjen Perhubungan Darat, 2009.

[13] D. PU, Data Jalan & Jembatan Selama Pelita IV, Departemen Pekerjaan Umum Dirjen Bina Marga, 2011.

[14] J. J. Siang, Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan Matlab 1st ed, Yogyakarta: Andi Yogyakarta, 2004.

[15] Dong, Y, Xia, Z, Tu dan M. &. Xing, An Optimization Method For Selecting Parameters In Support Vector Machines, Sixth International Conference On Machine Learning And Applications, 2007.