

# Predictive Model of Passengers Trans Metro Bandung Encouraging Smart Transportation

Nova Nurviana  
School of Electrical Engineering  
and Informatics  
Bandung Institute of Technology  
Bandung, Indonesia  
nova.nurviana@s.itb.ac.id

Sigit Kusumo Ati  
School of Electrical Engineering  
and Informatics  
Bandung Institute of Technology  
Bandung, Indonesia  
sigitati@gmail.com

Hani Purwati Hanifah  
School of Electrical Engineering  
and Informatics  
Bandung Institute of Technology  
Bandung, Indonesia  
hani.purwati@students.itb.ac.id

*Abstract—Smart city can be accomplished by merging all components to be smart system. One of significant components is transportation. Everybody moves, everybody need to reach other place safely, efficiency, and comfortably. In order to get smart, we have to analyze the best way to improve capability. One of the best ways is to make predictive model that one may conceive strategies attracting passengers. Using Bayesian dan SVM as comparison will produce best fit model for Trans Metro Bandung passenger data. Better strategy is an emergence needed to produce Trans Metro Bandung better performance because the model predicted there will be no significant growth for Trans Metro Bandung passenger.*

*Keywords—smart, transportation, Trans Metro Bandung, predictive, SVM, Bayesian*

## I. INTRODUCTION

Trans Metro Bandung adalah Bus Rapid Transit di kota Bandung yang diresmikan pada tanggal 22 Desember 2004. Saat ini Trans Metro Bandung memiliki beberapa trayek khusus, trayek tersebut terbagi menjadi 4 koridor utama, antara lain Koridor I: Cibiru – Elang, Koridor II: Cicaheum-Cibeureum, Koridor III: Cicaheum-Sarijadi. Koridor IV:Antapani-Terminal Leuwi Panjang. Masing- masing koridor tersebut memiliki sejumlah halte untuk menaikkan dan menurunkan penumpang.

Salah satu permasalahan yang dihadapi oleh TMB adalah jumlah penumpang yang sedikit. Dari permasalahan yang ada, kemudian dianalisis jumlah penumpang dengan menggunakan predictive model sehingga dalam perkembangan revolusi industri 4.0 teknologi pada transportasi akan mempermudah kehidupan [1].

Predictive model banyak digunakan peneliti sebagai evaluasi bagi penerapan sebuah teknologi atau kebijakan [2]. Selain itu, pada bidang tertentu, predictive model sangat dibutuhkan sebelum diaplikasikan sebuah penelitian, yaitu pada bidang medis dan kesehatan, dimana pengaplikasian obat atau teknologi akan bergantung pada hasil predictive model [3]. Pada bidang planologi, dimana predictive model sangat berperan penting dalam perencanaan wilayah. Misalnya, penelitian [4] yang mengklasifikasi wilayah sesuai dengan zona risiko bencananya dapat menjadi masukan bagi pemerintah untuk memutuskan apakah daerah tersebut layak menjadi pusat pemerintahan atau tidak [5]. Tentunya, pengklasifikasian tersebut dengan cara melihat struktur tanah, prediksi banjir, dan sejarah dari daerah tersebut apakah pernah terjadi bencana sebelumnya.

Pengaplikasian predictive model pada sarana transportasi telah diteliti oleh [6]-[12]. Dari kesemua penelitian, yang meneliti mengenai bis adalah [8] dimana metode yang digunakan bersifat mixed logical dinamik (MLD) antara motion dengan penumpang, sedangkan dalam paper ini hanya meneliti mengenai penumpang, sehingga studi literatur diteruskan hingga terdapat sebuah penelitian yang sama-sama meneliti penumpang dan memiliki tingkat akurasi yang tinggi, penelitian [13] ini memakai algoritma Bayes sebagai model prediksi. Sebagai bahan komparasi, dipakai Support Vector Machine dikarenakan algoritma ini umum dipakai sehingga banyak contoh-contoh program yang dapat dipakai untuk mengaplikasikan penelitian di Bus Trans Metro Bandung. Adapun peneliti yang memakai SVM sebagai predictive model adalah [14][15][16].

## II. METHOD/MATERIAL

### A. Support Vector Machine

Metode SVM (Support Vector Machine) merupakan metode untuk mengklasifikasikan dan meregresi suatu data. Awal mulanya SVM digunakan pada kasus yang dapat dipisahkan secara linear, yaitu dengan cara memilah data dan margin. Namun saat ini SVM dikembangkan pada kasus yang non linear, dengan menggunakan hyperplane untuk memaksimalkan jarak antar beberapa kelas. Metode yang digunakan oleh SVM adalah metode learning machine, yang dapat digunakan untuk pengenalan pola. SVM bekerja dengan prinsip Structural Risk Minimization yang bertujuan menemukan hyperplane terbaik yang memilah 2 buah class pada ruang input. Hyperplane digunakan sebagai pemilah dari class class yang berbeda, atau sering digambarkan sebagai sebuah garis yang memisahkan problem.

### B. Naive Bayes

Metode yang lain yang dapat digunakan untuk klasifikasi adalah Metode Naïve Bayes. Metode ini merupakan suatu metode probabilitas dan statistik untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa lalu. Ciri utamanya adalah asumsi yang sangat kuat (naïf) akan independensi dari masing– masing kondisi/kejadian. Kelebihan dari metode ini antara lain, dapat menangani kuantitatif dan data diskrit, selain itu naïve bayes kokoh untuk titik noise yang diisolasi, misalkan rata-rata titik dari hasil peluang bersyarat data. Untuk mengestimasi parameter yang dibutuhkan untuk klasifikasi, metode ini hanya memerlukan sejumlah kecil data pelatihan. Kelebihan yang lain adalah metode ini mampu menangani nilai yang hilang selama perhitungan estimasi peluang. Metode Naïve Bayes cepat, efisiensi dan kokoh terhadap atribut yang tidak relevan.

Prediksi Bayes didasarkan pada teorema Bayes dengan formula umum sebagai berikut :

$$P(H | E) = \frac{P(E | H) \times P(H)}{P(E)}$$

Penjelasan dari formula tersebut adalah sebagai berikut :

$P(H|E)$  : Probabilitas akhir bersyarat (conditional probability) suatu hipotesis H terjadi jika diberikan bukti (evidence) E terjadi.

$P(E|H)$  : Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H.

$P(H)$  : Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.

$P(E)$  : Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain.

### C. Data Mining

Data mining merupakan serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu database dengan melakukan penggalian pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam database. Tahapan dalam data mining antara lain:

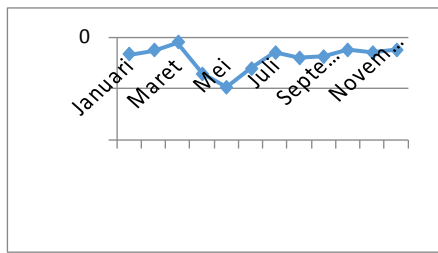
1. Data cleaning (untuk menghilangkan noise data yang tidak konsisten) Data integration (di mana sumber data yang terpecah dapat disatukan).
2. Data selection (di mana data yang relevan dengan tugas analisis dikembalikan ke dalam database).
3. Data transformation (di mana data berubah atau bersatu menjadi bentuk yang tepat untuk menambang dengan ringkasan performa atau operasi agresif).

4. Knowledge Discovery (proses esensial di mana metode yang intelejen digunakan untuk mengekstrak pola data)
5. Pattern evolution (untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan atas beberapa tindakan yang menarik).
6. Knowledge presentation (di mana gambaran teknik visualisasi dan pengetahuan digunakan untuk memberikan pengetahuan yang telah ditambang kepada user).

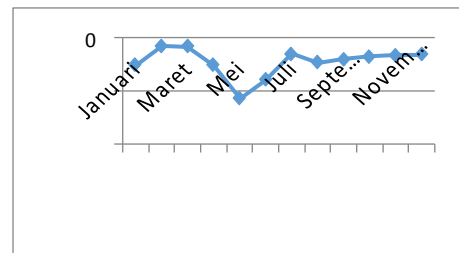
### III. ANALISIS

Data yang akan diolah adalah data di tahun 2018 yang berada di koridor 1,2,3 dan 4. Sebelum melakukan analisa terlebih dahulu dilakukan pengolahan (preprocessing) data. Data yang dimiliki adalah data aktivitas Bus Trans Metro Bandung masing-masing koridor, kemudian data tersebut digabungkan berdasarkan field yang sejenis. Dari hasil penggabungan data, dilakukan proses cleaning untuk menghilangkan atribut yang tidak relevan dan menghilang row dengan nilai null dan row pencilan (noise tidak konsisten). Melalui proses selection, dipilih field yang meliputi: Tanggal, Hari, Koridor, Jumlah Penumpang (berdasarkan kelompok pelajar, umum dan total penumpang dan rata-rata penumpang), Jumlah Armada dan Kapasitas Bus.

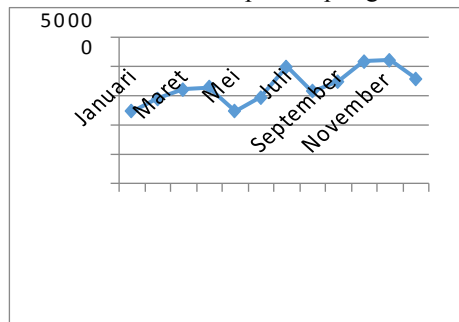
Dari data tersebut dapat dilakukan analisa deskriptif jumlah penumpang per-bulan, per-koridor tahun 2018, ditunjukkan pada gambar berikut



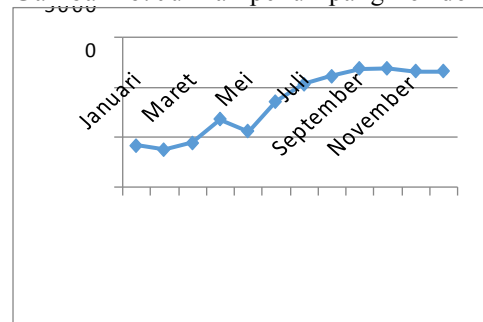
Gambar 1a. Jumlah penumpang koridor 1



Gambar 1b. Jumlah penumpang koridor2



Gambar 1c. Jumlah Penumpang Koridor3



Gambar 1d. Jumlah Penumpang Koridor4

Dari gambar 1a,1b,1c,1d dapat dilihat bahwa pada bulan mei rata-rata mengalami penurunan jumlah penumpang. Selain itu dari keempat koridor tersebut, koridor keempat yang memiliki jumlah penumpang paling sedikit.

Preprocessing data selanjutnya dilakukan dengan melakukan transformasi data melalui penentuan standard range dari data penumpang pelajar dan umum, yang meliputi :

Table1. Standard jumlah penumpang umum dan pelajar

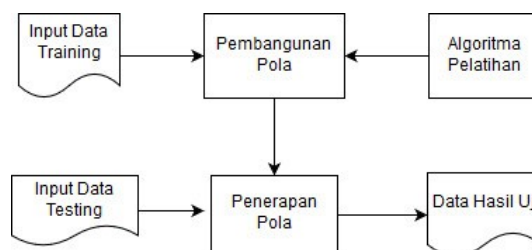
Standard penumpang	Jumlah Penumpang
Range 1 penumpang pelajar	0-100
Range 2 penumpang pelajar	101-300
Range 3 penumpang pelajar	301-600
Range 4 penumpang pelajar	> 600
Range 1 penumpang umum	< 300
Range 2 penumpang umum	301-600
Range 3 penumpang umum	601-1000
Range 4 penumpang umum	>1000

Dengan hasil sebagai berikut :

HARI	PELAJAR	UMUM	KORIDOR	RANGE_HARI	RANGE_PELAJAR	RANGE_UMUM	KORIDOR
Monday	66.00	447.00	1	1	1	1	1
Tuesday	58.00	282.00	1	2	1	1	1
Wednesday	90.00	540.00	1	3	1	1	1
Thursday	132.00	552.00	1	4	2	1	1
Friday	126.00	619.00	1	5	2	1	1
Saturday	124.00	609.00	1	6	2	1	1
Sunday	108.00	608.00	1	7	2	1	1

Gambar 2. Standar format dalam range

Data hasil proses transformasi tersebut kemudian dilakukan proses pengecekan kekonsistenan data. Selanjutnya dilakukan penentuan pola dengan menggunakan metode naive bayes. Tahapan dalam melakukan pengklasifikasian dengan menggunakan metode naive bayes ini adalah sebagai berikut:



Gambar 3. Tahapan pengklasifikasian

Penentuan pola dilakukan melalui suatu persamaan probability, nilai probability yang terbesar itulah yang akan diambil. Berdasarkan rumus naive bayes :

$$P(H | E) = \frac{P(E | H) \times P(H)}{P(E)}$$

Jika E adalah vektor masukan, misal data jumlah penumpang pelajar, dan H adalah label kelas, yaitu Range 1,2,3,4. Maka Naive Bayes dituliskan dengan P(H|E). Notasi tersebut berarti probabilitas label kelas H didapatkan setelah fitur-fitur E diamati. Notasi ini disebut juga probabilitas akhir (posterior probability) untuk H, sedangkan P(H) disebut probabilitas awal (prior probability) H. Model dari kombinasi E dan H dari data training dilakukan proses learning probabilitas P(H|E). Setelah model terbentuk, data Testing E' dapat dilakukan klasifikasi dengan mencari nilai H' dengan memaksimalkan nilai P(E'|H') yang didapat.

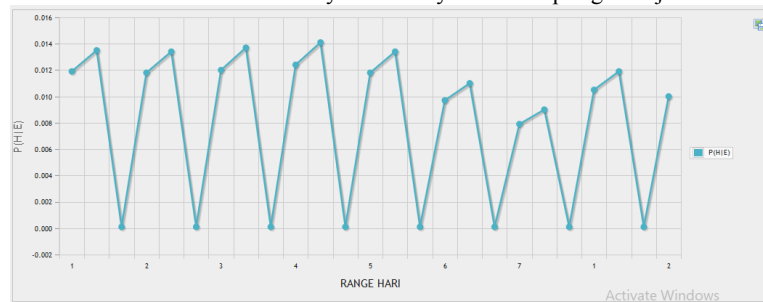
Table2. Nilai P(H) kelas penumpang

Penumpang	Kelas	P(H)
Pelajar	Range1	0.46
Pelajar	Range2	0.49
Pelajar	Range3	0.05
Pelajar	Range4	0
Umum	Range1	0.78
Umum	Range2	0.01
Umum	Range3	0.17
Umum	Range4	0.04

$P(E|H)$  merupakan nilai probabilitas Hari tertentu dan Koridor tertentu pada kelas penumpang Range tertentu. Digambarkan dengan tabel sebagai berikut :

RANGE HARI	KORIDOR	RANGE PELAJAR	P(H)	P(E)	P(E H)	P(H E)
1	1	1	0.46	0.056293805	0.025895151	0.0119
		2	0.49	0.056293805	0.027583965	0.0135
		3	0.05	0.056293805	0.00281469	0.0001
2	1	1	0.46	0.055979824	0.025750719	0.0118
		2	0.49	0.055979824	0.027430114	0.0134
		3	0.05	0.055979824	0.002798991	0.0001
3	1	1	0.46	0.056866062	0.026158388	0.0120
		2	0.49	0.056866062	0.02786437	0.0137
		3	0.05	0.056866062	0.002843303	0.0001
4	1	1	0.46	0.058623344	0.026966738	0.0124
		2	0.49	0.058623344	0.028725439	0.0141
		3	0.05	0.058623344	0.002931167	0.0001
5	1	1	0.46	0.055726613	0.025634242	0.0118
		2	0.49	0.055726613	0.027306041	0.0134
		3	0.05	0.055726613	0.002786331	0.0001
6	1	1	0.46	0.045831139	0.021082324	0.0097
		2	0.49	0.045831139	0.022457258	0.0110
		3	0.05	0.045831139	0.002291557	0.0001
7	1	1	0.46	0.037515699	0.017257222	0.0079
		2	0.49	0.037515699	0.018382693	0.0090
		3	0.05	0.037515699	0.001875785	0.0001

Gambar 4. Probability naive bayes Penumpang Pelajar



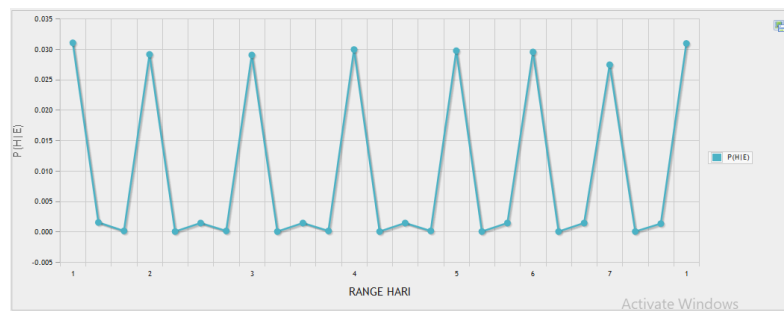
Gambar 5. Grafik probabilitly naive bayes penumpang pelajar per hari

Dari hasil penghitungan tersebut, dapat disimpulkan bahwa dari hari senin s/d minggu rata-rata penumpang pelajar di koridor 1 berada di range 2. Sehingga diprediksi jumlah penumpang di tahun yang akan datang di koridor 1 adalah antara 101 - 300 orang per harinya. Setelah dilakukan penghitungan lebih lanjut di koridor 2 dan 3 diprediksi bahwa jumlah penumpang yang akan datang juga berada di range 2 juga yaitu di kisaran 101 - 300 penumpang tiap harinya.

Dilakukan kembali penghitungan probabilitas naive bayes untuk penumpang Umum, dengan hasil sebagai berikut:

RANGE HARI	KORIDOR	RANGE UMUM	P(H)	P(E)	P(E H)	P(H E)
1	1	1	0.78	0.050930247	0.039725593	0.0310
		3	0.17	0.050930247	0.008658142	0.0015
		4	0.04	0.050930247	0.00203721	0.0001
2	1	1	0.78	0.047798545	0.037282865	0.0291
		2	0.01	0.047798545	0.000477985	0.0000
		3	0.17	0.047798545	0.008125753	0.0014
		4	0.04	0.047798545	0.001911942	0.0001
3	1	1	0.78	0.047732633	0.037231454	0.0290
		2	0.01	0.047732633	0.000477326	0.0000
		3	0.17	0.047732633	0.008114548	0.0014
		4	0.04	0.047732633	0.001909305	0.0001
4	1	1	0.78	0.04909006	0.038290247	0.0299
		2	0.01	0.04909006	0.000490901	0.0000
		3	0.17	0.04909006	0.00834531	0.0014
		4	0.04	0.04909006	0.001963602	0.0001
5	1	1	0.78	0.048778315	0.038047086	0.0297
		2	0.01	0.048778315	0.000487783	0.0000
		3	0.17	0.048778315	0.008292314	0.0014
6	1	1	0.78	0.048562766	0.037878957	0.0295
		2	0.01	0.048562766	0.000485628	0.0000
		3	0.17	0.048562766	0.00825567	0.0014

Gambar 6. Probability Naive Bayes Penumpang Umum



Gambar 7. Grafik probabilitly naive bayes penumpang Umum per hari

Pada Koridor 1 hari Senin s/d Minggu rata-rata jumlah penumpang berada pada range 1 yaitu kurang dari 300 penumpang. Setelah dilakukan penghitungan kembali jumlah penumpang pada koridor 2,3, dan 4 juga berada di range1. Sehingga dapat diprediksi pada masa yang akan datang (tahun depan), jumlah penumpang umum bus TMB per harinya berkisar 1-300 orang per hari.

Data tersebut dapat pula diklasifikasikan menggunakan SVM, dengan hasil sebagai berikut :

Row No.	umum	prediction(u...	confidence(...	confidence(...	confidence(...	hari	koridor
1	range2	range1	1	0	0	3	1
2	range2	range1	1	0	0	4	1
3	range2	range1	1	0	0	5	1
4	range2	range1	1	0	0	6	1
5	range2	range1	1	0	0	7	1
6	range3	range1	1	0	0	4	1
7	range2	range1	1	0	0	5	1
8	range2	range1	1	0	0	6	1
9	range3	range1	1	0	0	4	1
10	range3	range1	1	0	0	2	1

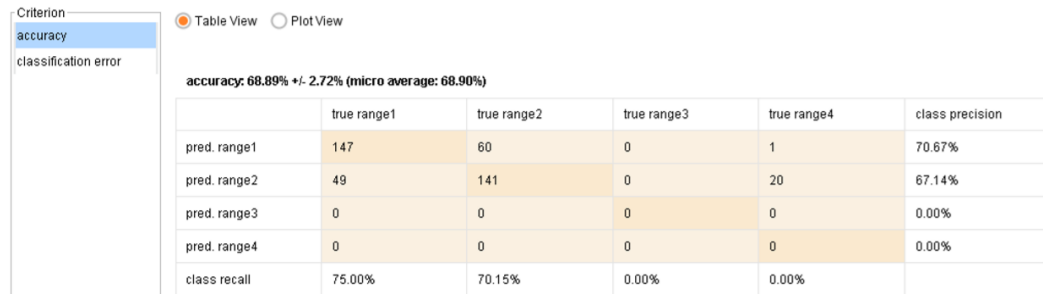
Gambar 6. SVM klasifikasi penumpang Umum

Pada penumpang umum, menunjukkan hasil yang sama ketika menggunakan metode Naive bayes, yaitu rata-rata penumpang umum berada di range 1. Sehingga dengan

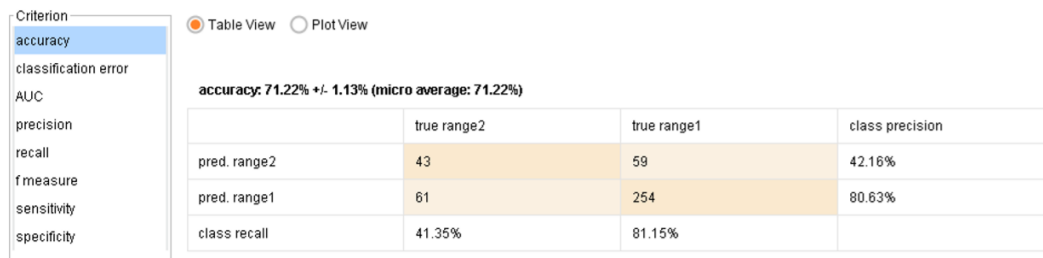
metode SVM juga akan menghasilkan hasil prediksi yang sama dengan ketika menggunakan metode Naive Bayes, yaitu pada tahun yang akan datang jumlah penumpang umum di koridor 1,2,3 dan 4 adalah di kisaran 1-300 orang per hari. Perbedaan antara SVM dan Naive Bayes adalah pada tingkat keakurasiannya.

#### IV. EVALUASI MODEL

Evaluasi diperlukan untuk mengetahui keakuratan model naive bayes, ditunjukkan dalam gambar berikut:



Gambar 7. Evaluasi Model prediksi penumpang pelajar dalam 4 range kelas metode SVM



Gambar 8. Evaluasi Model prediksi penumpang umum dalam 4 range kelas metode Naive bayes

Dari gambar tersebut dapat diketahui bahwa model naive bayes dengan pengelompokan 4 range kelas, pada penumpang pelajar nilai keakurasiannya adalah 68,8%, sedangkan pada penumpang umum adalah 71%.

#### V. CONCLUSION

Dalam melakukan analisa terhadap data Bus Trans Metro Bandung, maka dapat disimpulkan bahwa:

1. Dari hasil analisa deskriptif jumlah penumpang perkoridor tidak menunjukkan peningkatan yang signifikan di tahun 2018. Setiap armada bus hanya terisi 10-20% dari kapasitas bus.
2. Koridor 4 memiliki jumlah penumpang paling sedikit, Hal ini perlu dianalisa lebih lanjut penyebabnya, terkait dengan perutean di koridor 4, detinasi orang berkunjung di jalur koridor, misal: kantor, sekolah, dll.
3. Berdasarkan pola klasifikasi dari metode naive bayes, dapat diklasifikasi jumlah penumpang bus yang akan datang untuk penumpang umum adalah berada di kisaran 1-300 orang perharinya dan untuk pelajar berada di kisaran 100 - 300 orang perharinya.

## REFERENCES

- [1] Faruqi, Umar Al. Survey Paper: Future Service in Industry 5.0. Published in Jurnal Sistem Cerdas 2019 Volume 02 No 01 ISSN : 2622-8254 Hal : 67 - 79
- [2] Hovsepyan, Aram, et al. Is Newer Always Better? The Case of Vulnerability Prediction Model. Published in Proceeding ESEM '16 Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement Article No. 26
- [3] Peng T, et al. Predictive Modeling of Drug Effects on Electrocardiograms. Published in Journal Computers in Biology and Medicine 108 (2019) 332-344
- [4] Kempf, Michael. The Application of GIS and Satellite Imagery in Archaeological Land-Use Reconstruction: A Predictive Model? Published in Journal of Archaeological Science: Report
- [5] Schatz, Edward. When Capital Cities Move: The Political Geographical of Nation and State Building
- [6] Liu, Chunxia, et al. An Improved Grey Neural Network Model for Predicting Transportation Disruptions. Published in Journal Expert Systems with Application 45(2016) 331-340
- [7] Zhou, Huide, et al. An Approach of Model Predictive Control for Urban Transportation Network. Published in 13th IFAC Symposium on Large Scale Complex Systems: Theory and Applications July 7-10, 2013. Shanghai, China
- [8] I.I Sirmatel and N. Geroliminis. Dynamical Modeling and Predictive Control of Bus Transport System: A Hybrid Systems Approach. Published in IFAC Papers OnLine 50-1 (2017) 7499-7504
- [9] Zhou, Zhao, et al. A Congestion Eliminating Control Method for Large-Scale Urban Traffic Network. Published in 13th IFAC Symposium on Large Scale Complex Systems: Theory and Applications July 7-10, 2013. Shanghai, China
- [10] Khanmohamadi, Masoud, et al. A Security Vulnerability Analysis Model for Dangerous Goods Transportation by Rail – Case Study: Chlorine Transportation in Texas-Illinois. Published in Journal Safety Science 110 (2018) 230-241
- [11] Wu, Yizheng, et al. Modeling Health Equity in Active Transportation Planning. Published in Journal Transportation Research Part D 67 (2019) 528-540
- [12] Huang, Dengpeng, et al. Modelling of Serrated Chip Formation Processes Using the Stabilized Optimal Transportation Meshfree Method. Published in International Journal of Mechanical Science Volume 155, May 2019, Pages 323-333
- [13] Sona Taheri, Musa Mammadov, "Learning The Naive Bayes Classifier With Optimization Model", Int. J. Appl. Math. Comput. Sci., 2013, Vol. 23, No. 4, 787-795, DOI: 10.2478/amcs-2013-0059
- [14] Feng, Yating, et al. Short Term Load Forecasting of Offshore Oil Field Microgrids Based on DA-SVM. Published in 10th International Conference on Applied Energy (ICAE2018), 22-25 August 2018, Hong Kong, China
- [15] Dhamecha, Tejas Indulal, et al. Between-Subclass Piece-wise Linear Solutions in Large Scale Kernel SVM Learning. Published in Journal Pattern Recognition.
- [16] Chao, Luo, et al. A Novel Reconstructed Training-Set SVM with Roulette Cooperative Coevolution for Financial Time Series Classification. Published in Journal Experts Systems with Applications 123(2019)283-298
- [17] Alfa Shaleh, 'Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga', Citec Journal Vol.2
- [18] Babajide O. Afeni, Thomas I. Arulebal, Iyanuoluwa A. Oloyede, "Hypertension Prediction System Using Naive Bayes Classifier", Journal of Advances in Mathematics and Computer Science, Previously known as British Journal of Mathematics & Computer Science