ESTIMASI TUJUAN PENUMPANG MENGGUNAKAN *PREDICTIVE MODEL* DENGAN DATA *SMART CARD*

Inkreswari Retno Hardini[1], Goklas Henry Agus Panjaitan[2], Ekky Novriza Alam[3], Sita Rosida[4]
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung Bandung,
Indonesia
inkreswari@gmail.com[1], goklasif10029@gmail.com[2] , 23517006@std.stei.itb.ac.id[3],
23517025@std.stei.itb.ac.id[4]

Abstract—Bus Rapid Transit (BRT) is one of the main choices of public transportation that supports mobility of Jakarta community. As one of the main choices of public transportation, BRT should provide good service and always improve its performance. Needs for moving or mobility will cause a problem if the moving itself is heading at the same area and at the same time. That will cause some problems which are often faced in urban areas such as traffic and delay. To overcome those problems there needs to be a strategy to build good public transportation planning, besides need to know individual travel patterns to overcome problems and improve BRT service. In case to realize those plans needs to be built origin-destination (O-D) matrix. O-D matrix is a matrix that each cell is an amount of trip from the source(row) to the destination (column). O-D matrix is beneficial for analysis, design and public transportation management. O-D matrix also provides useful information like amount of trip between 2 different locations, that can be utilized as fundamental information for decision making for three levels of strategic management (long term planning), tactic (service adjustment and network development), and operational (scheduling, passenger statistic, and performance indicator). To build O-D matrix is required a predictive model that can be measured to predict passenger destination. The predictive model will be build using classification algorithms such as Decision Tree and K- Nearest Neighbor (KNN).

Keywords—predictive model, classification, *decision tree*, *k-nearest neighbor*, *bus rapid transit*.

I. PENDAHULUAN

Kebutuhan akan pergerakan akan mengakibatkan permasalahan, apabila terjadi pergerakan untuk tujuan yang sama disuatu daerah dan dalam waktu yang bersamaan, tentunya akan menyebabkan permasalahan yaitu kemacetan dan keterlambatan. Bus rapid transit merupakan salah satu transportasi umum yang menunjang aktivitas pergerakan masyarakat Jakarta. Oleh karenanya bus rapid transit (BRT) harus dapat memberikan layanan yang baik dan senantiasa memperbaiki kinerja. Layanan yang baik dapat diukur dari berbagai macam indikator, beberapa diantaranya adalah waktu tunggu penumpang di halte, tingkat kepadatan penumpang didalam sebuah bus yang harus sesuai dengan kapasitas, dan kemudahan pengumpulan tarif perjalanan.

Untuk meningkatkan layanan BRT serta untuk mengatasi permasalahan yang terjadi seperti kemacetan dan keterlambatan, maka harus diketahui pola jumlah kebutuhan transportasi dan pola individu penumpang. Informasi mengenai kedua hal tersebut dapat diperoleh melalui matriks *origin-destination*. Matriks *origin-destination* adalah matriks yang setiap selnya merupakan jumlah perjalanan dari titik asal (baris) ke tujuan (kolom) [1][2].

Untuk meningkatkan kualitas dan keefektifan transportasi umum, diterapkanlah *automated fare collection* (AFC), atau dikenal sebagai transit *smart card* (SC). Data transaksi yang terekam melalui AFC memungkinkan untuk memberikan pengetahuan atau *insight* baru yang bermanfaat bagi ilmu pengetahuan, tetapi diperlukan metodologi untuk mengekstrak pengetahuan tersebut karena kebanyakan sistem AFC awalnya tidak dirancang untuk pengumpulan data [3].

Data transaksi *smart card* melalui *Automated fare collection* (AFC) sistem menjadi sumber informasi utama untuk membuat matriks *origin-destination* (O-D) transportasi

©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019

umum. Dengan adanya sumber informasi tersebut, memungkinkan pemahaman yang lebih baik tentang pola perjalanan individu dan perencanaan strategis terkait transportasi umum kualitasnya semakin meningkat karena ditunjang data yang *real*. Matriks O-D penting untuk analisis, desain, dan manajemen transportasi umum. Selain itu informasi yang terkandung adalah mengenai jumlah perjalanan antara dua lokasi yang berbeda, yang dapat dimanfaatkan sebagai informasi dasar pengambil keputusan untuk tiga tingkatan manajemen strategis, taktis, dan operasional [3][4][5][6][7].

Pada masa awal beroperasi, BRT Transjakarta menggunakan sistem *entry-only*. Lalu bertambahnya waktu, bus rapid transit Transjakarta mulai menggunakan sistem yang merekam transaksi asal dan tujuan penumpang setiap harinya, yang bernama AFC *entry-exit*. Namun, data transaksi yang terekam oleh system AFC tidak dapat merepresentasikan keseluruhan data penumpang asal dan tujuan karena kendala infrastruktur yang masih belum sepenuhnya berfungsi dan perilaku penumpang yang tidak taat [8].

Mengingat system AFC yang digunakan serta sumber data yang dimiliki keduanya berbeda, maka diperlukanlah proses pengolahan data yang berbeda pula untuk membentuk matriks O-D. Terdapat penelitian yang menggunakan model *Trip-chaining* yang merupakan metode yang paling banyak digunakan untuk melakukan estimasi tujuan penumpang pada system *entry-only* untuk membuat matriks O-D [6]. *Trip-chaining* model pertama kali diterapkan di New York untuk keperluan menyimpulkan tujuan penumpang. Hal tersebut dilakukan dengan menggunakan dua asumsi utama yaitu stasiun tujuan dari perjalanan saat ini merupakan stasiun asal dari perjalanan berikutnya dan sebagian besar pengguna transportasi umum mengakhiri perjalanannya pada stasiun mereka memulai perjalanan pada hari tersebut [9]. Kemudian dalam penelitian yang dilakukan oleh Zhao terdapat asumsi tambahan bahwa penumpang menggunakan jenis transportasi yang sama diantara dua perjalanan tersebut [10].

Terdapat penelitian yang menggunakan analisis big data untuk mengestimasi matriks O-D BRT Transjakarta dengan menggunakan data *entry-only* dan model *trip-chaining*. Penelitian tersebut berhasil membangun matriks O-D tetapi 36,5% transaksi penumpang tidak dapat diprediksi [8]. Oleh karena itu diperlukan *predictive model* yang dapat diukur untuk melakukan prediksi tujuan penumpang yang dapat diukur untuk melakukan prediksi tujuan penumpang yang dapat dimanfaatkan untuk membangun matriks O-D.

Predictive model dapat dilakukan dengan menggunakan algoritma klasifikasi seperti *decision tree* yang telah dilakukan dalam penelitian [11] untuk prediksi tujuan pengemudi dan pernah digunakan juga untuk prediksi yang menggunakan *mobile data* [12]. Algoritma klasifikasi lainnya yang dapat digunakan yaitu K-Nearests Neighbor yang pernah digunakan dalam penelitian [13][14] untuk melakukan prediksi tujuan dan rute kapal dengan menggunakan data Automatic Identification System (AIS). Algoritma klasifikasi ini dapat menjadi alternative untuk membangun *predictive model* dengan data transaksi *smart card* BRT.

Melalui prediksi estimasi tujuan penumpang yang akurat, diharapkan dapat membantu dalam pembentukan perencanaan strategis sehingga dapat mengurangi keterlambatan operasional BRT, kepadatan penumpang dan kemudahan pengumpulan tarif penumpang. Konsep ini peningkatan layanan menjadi lebih efektif dan efisien ini tentunya sejalan dengan konsep future service di dunia industri [15].

II. METODE

A. Rancangan Solusi

Untuk membentuk matriks *origin-destination* menggunakan data transaksi *smart card* penumpang transportasi umum diperlukan informasi asal dan tujuan penumpang. Sebagian besar sistem AFC merekam informasi asal perjalanan tetapi tidak merekam informasi tujuan perjalanan penumpang. Dari batasan yang diberikan oleh penerapan sistem AFC, model *trip-chaining* merupakan metode yang paling sering digunakan untuk melakukan

©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019

prediksi tujuan perjalanan penumpang. Namun hasil prediksi tujuan penumpang yang dihasilkan oleh model *trip-chaining* akan sulit untuk melakukan pengujian tanpa data yang cukup dan memadai.

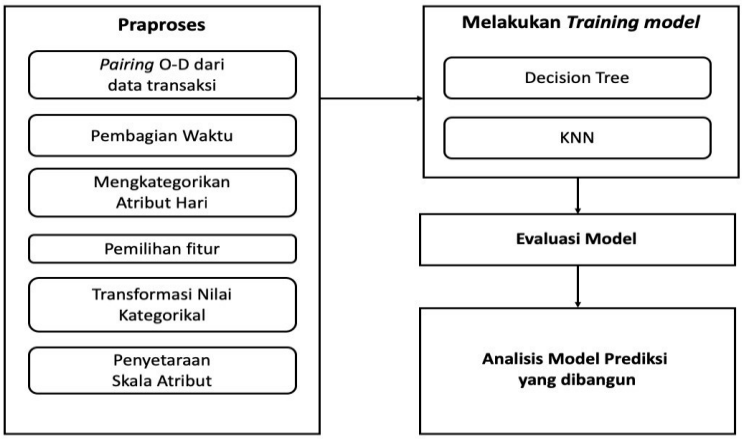
Dengan menggunakan data dari sistem AFC *entry-exit* dari sumber data, data asal dan tujuan perjalanan penumpang dijadikan sebagai data yang sudah berlabel untuk membangun model predisksi terhadap tujuan penumpang dengan menggunakan algoritma *decision tree* dan KNN. Dua algoritma ini digunakan karena berhasil digunakan oleh penelitian terdahulu untuk memprediksi tujuan pada domain yang berbeda [13][12][14].

Berdasarkan data yang dimiliki dan pengaturan rute transportasi umum, prediksi terhadap tujuan dapat dikelompokkan menjadi tiga, yaitu halte tujuan, sub-koridor tujuan, dan koridor tujuan. Halte tujuan penumpang diprediksi untuk mengetahui secara detail O-D dari suatu halte ke halte lain. Sub-koridor tujuan penumpang diprediksi untuk mengetahui O-D dari bagian koridor. Sedangkan koridor tujuan diketahui untuk mengetahui jumlah transit antar koridor.

Untuk membangun *predictive model* maka data asal dan tujuan penumpang akan dibagi menjadi *training set* dan *testing set*. Kedua model ini akan digunakan untuk kemungkinan solusi dalam membentuk matriks *origin-destination* yang dapar diuji. Hasil algoritma ini dibandingkan untuk membentuk matriks *origin-destination*.

1) Tahapan Solusi

Untuk melakukan prediksi tujuan perjalanan penumpang dengan menggunakan data transaksi *smart card*, maka dilakukan 4 tahapan utama yaitu praproses, *training model*, evaluasi model dan analisis model prediksi yang ditunjukkan pada Gambar *1*.



Gambar 1. Tahapan Solusi

2) Evaluasi Rancangan Solusi

Evaluasi model yang dibangun untuk memprediksi tujuan penumpang, dilakukan dengan pengujian sisi akurasi terhadap data uji. Pengukuran akurasi menggunakan matriks akurasi. Model pengukuran menggunakan matriks akurasi dipilih karena hasil prediksi berupa klasifikasi atau prediksi benar atau salah. Untuk dapat membangun matriks akurasi, maka perlu dihitung jumlah tujuan penumpang yang benar terhadap data uji. Hasil prediksi benar (*true positive*), jumlah transaksi yang salah diprediksi (*false positive*), jumlah transaksi yang gagal (*false negative*).

Setelah nilai matriks akurasi didapatkan maka dapat dihitung *precision* dan *recall*. Kedua pengukuran ini cukup baik memberikan hasil terhadap model prediksi. *Precision* merepresentasikan tingkat ketepatan model untuk memprediksi tujuan perjalanan penumpang. Sedangkan *recall* merepresentasikan kemampuan model untuk memprediksi seluruh tujuan perjalanan penumpang.

©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019

Untuk menentukan akurasi dari prediksi berdasarkan kedua nilai (*precision* dan *recall*) yang merepresentasikan makna yang berbeda, maka dilakukan perhitungan *f-measure* untuk menggabungkan kedua nilai tersebut (*precision* dan *recall*) dan didefinisikan sebagai pengukuran tunggal

B. Pengumpulan Data

1) Sumber Data

Sumber data yang digunakan berasal dari data transaksi *smart card* yang didapatkan dari Badan Usaha Bus Rapid Transit Jakarta.

2) Deskripsi Data

Data yang digunakan merupakan data transaksi *smart card* penumpang BRT Transjakarta yang direkam melalui sistem AFC. Sistem AFC yang diterapkan oleh BRT Transjakarta menggunakan sistem AFC entry-exit yang merekam informasi asal dan tujuan penumpang. Setiap kali penumpang keluar masuk halte akan direkam sebagai satu transaksi. Informasi yang disimpan setiap kali transaksi yaitu timestamp, id kartu, bank penerbit kartu, koridor, subkoridor, halte, dan jenis transaksi keluar atau masuk halte. Dari data diketahui terdapat 13 koridor yang terbagi menjadi 43 sub-koridor. Dalam 43 sub- koridor terdapat total 233 halte. Data transaksi *smart card* mencakup atribut-atribut yang tertera pada Tabel *1*.

Tabel 1. Deskripsi Data

Atribut	Deskripsi
tanggal_jam	Tanggal dan jam penumpang melakukan transaksi
id_kartu	Nomor kartu
kartu	Jenis kartu yang digunakan
koridor	Jalur/koridor BRT asal perjalanan penumpang
subkoridor	pembagian zona koridor asal perjalanan penumpang
halte	Halte asal penumpang
flag	In/out

Jumlah data yang didapatkan sebanyak 16.099.914 data dan setelah melakukan praproses data, maka data yang digunakan sebanyak 7.895.330 data yang diambil dari pasangan *in-out* dan dengan maksimal jarak transaksi selama 240 menit.

3) Praproses Data

Seperti yang telah dijelaskan pada tahapan solusi, praproses data dilakukan untuk mempersiapkan data sumber untuk dapat diproses untuk membangun model. Berdasarkan hasil analisa metode, praposes yang dilakukan adalah pengelompokan data, *pairing* O-D dari data transaksi, menentukan *training set* dan *testing set*, serta pelabelan outlier.

- *Pairing* O-D dari data transaksi
  Pairing dari data transaksi penumpang *tap-in* dan *tap-out*. Data transaksi yang terlihat pada Tabel *2* tidak memperlihatkan keterhubungan antara asal dan tujuan dari setiap penumpang. Untuk itu dilakukan *pairing* antara *tap-in* dan *tap-out*. Hal ini dilakukan dengan tujuan mengetahui pasanfan (asal dan tujuan) dari setiap transaksi.
  Pairing data dilakukan berdasarkan nomor id kartu yang sama dan jarak waktu antara transaksi. Data terlebih dahulu diurutkan berdasarkan id kartu dan timestamp transaksi. Untuk setiap *tap-in* dilakukan pencarian transaksi *tap-out*
  yang memiliki id kartu yang sama dan juga jarak waktu terdekat.Tabel *3* memuat contoh data hasil *pairing* O-D.

- Pembagian waktu menjadi beberapa grup

©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019

Pengelompokan berdasarkan zona waktu dilakukan untuk memperoleh korelasi temporal transaksi pengguna transportasi umum secara efektif. Pembangunan zona waktu dilakukan dengan metode *data-driven* yaitu melihat kecenderungan kepadatan penumpang dan mendefinisikan zona waktu. Metode ini dilakukan agar pembagian zona waktu dapat merepresentasikan kebiasaan pengguna transportasi BRT.

Pengelompokan berdasarkan zona waktu dilakukan untuk memperoleh korelasi temporal transaksi pengguna transportasi umum secara efektif, pada penelitian ini 24 jam dalam sehari dibagi menjadi 12 periode waktu dengan jarak masing-masing 2 jam.

- 00.00 – 02.00
- 02.00 – 04.00
- 04.00 – 06.00
- 06.00 – 08.00
- 08.00 – 10.00
- 10.00 – 12.00
- 12.00 – 14.00
- 14.00 – 16.00
- 16.00 – 18.00
- 18.00 – 20.00
- 20.00 – 22.00
- 22.00 – 24.00

- Mengkategorikan atribut hari

  Kategori hari dibagi menjadi dua, yaitu hari libur dan hari kerja. Pengelompokan dilakukan dengan melakukan pendefinisian manual terhadap tanggal libur. Pengelompokan ini dilakukan karena kecenderungan tujuan penumpang pada hari kerja dan hari libur berbeda. Pada hari kerja, kecenderungan tujuan penumpang adalah halte dikawasan perkantoran.

- Pemilihan fitur

  Pemilihan fitur bertujuan untuk memilih fitur yang berpengaruh dan mengesampingkan fitur yang tidak berpengaruh dalam suatu kegiatan pemodelan atau penganalisaan data. Ada banyak alternatif yang bisa digunakan dan harus dilakukan secara *iterative* untuk menemukan fitur yang cocok.

  Dalam penelitian ini dipilih beberapa fitur yang berpengaruh untuk melakukan prediksi halte tujuan perjalanan. Fitur-fitur tersebut tercantum pada Tabel *2*.

Tabel 2. Fitur-fitur yang digunakan untuk membangun model prediksi tujuan

Fitur	Deskripsi	Information Gain
id_kartu	Merupakan identitas unik penumpang	0.422897
in_halte	Halte asal penumpang	0.147459
in_subkoridor	Pembagian zona koridor asal perjalanan penumpang	0.134915
in_koridor	Jalur/koridor BRT asal perjalanan penumpang	0.143483
time_zone	Pembagian waktu menjadi 12 zona waktu	0.140807
day_category	Hari kerja/hari libur	0.01044

©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019

Tabel 3. Contoh Data O-D dari hasil pairing

No	Card_ID	Kartu	In_Datetime	In_Halte	In_Subkoridor	In_Koridor	Out_Datetime	Out_Halte	Out_Subkoridor	Out_Koridor
1	145060684827	BCA	1/1/2018 0:01	Makro	K-07 South	[7] Kampung Rambutan – Kampung Melayu	1/1/2018 0:36	Bidara Cina	K-07 North	[7] Kampung Rambutan – Kampung Melayu
2	603293938205	MANDIRI	1/1/2018 0:01	BNN	K-07 Middle	[7] Kampung Rambutan – Kampung Melayu	1/1/2018 0:54	Pesakih	K-03 Outer	[3] Kalideres – Harmoni – Pasar Baru
3	145003554812	BCA	1/1/2018 0:01	Central Park arah Pluit	K-09 Inner East	[8] Lebak Bulus – Harmoni	1/1/2018 0:18	Penjaringan	K-09 Outer West	[9] Pinang Ranti - Pluit
4	145000869736	BCA	1/1/2018 0:01	Cempaka Timur	K-10 Middle North	[2] Pulo Gadung 1 – Harmoni	1/1/2018 1:02	Cempaka Timur	K-10 Middle North	[2] Pulo Gadung 1 – Harmoni
5	145000926618	BCA	1/1/2018 0:01	Cempaka Timur	K-10 Middle North	[2] Pulo Gadung 1 – Harmoni	1/1/2018 1:02	Cempaka Timur	K-10 Middle North	[2] Pulo Gadung 1 – Harmoni

III. HASIL DAN DISKUSI

A. Analisis Data

Data yang digunakan adalah data transaksi tap-in & tap-out *smart card* penumpang BRT Transjakarta. Sistem perekamanan menggunakan sistem AFC yang telah dijelaskan pada Bab sebelumnya. Informasi didapatkan setiap kali penumpang melakukan tap/transaksi. Informasi yang didapatkan yaitu timestamp, id kartu, bank penerbit kartu, koridor, subkoridor, halte, dan jenis transaksi masuk atau keluar halte. Contoh data dapat dilihat di Tabel *4*.

Tabel 4. Contoh Data *Smart card*

Tanggal_Jam	Pan	Kartu	Koridor	Subkoridor	Halte	Flag
1/23.2018 9:46	145000100070412	BCA	[1] Blok M - Kota	K-01 South	Masjid Agung	In
1/23.2018 10:33	145000100070412	BCA	[2] Pulo Gadung 1 - Harmoni	K-02 Inner North	Gambir 2	Out
1/19/2018 13:08	145000100058904	BCA	[1] Blok M - Kota	K-01 North	Olimo	In
1/19/2018 13:47	145000100058904	BCA	[1] Blok M - Kota	K-01 North	Stasiun Kota	Out

©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019

B. Pembangunan Model yang diusulkan

Pembangunan ini dapat menggunakan dua algoritma, yaitu KNN dan *decision tree*. Training model dilakukan dengan tiga target label prediksi, yaitu halte tujuan penumpang, sub-koridor tujuan penumpang dan koridor tujuan penumpang.

1) K-Nearest Neighbor

Algoritma KNN memiliki nilai K yang sangat berpengaruh dalam menentukan hasil klasifikasi. Nilai K pada kasus ini akan menentukan jumlah perjalanan yang mirip berdasarkan kedekatannya. Nilai optimum K akan didapatkan dengan melakukan percobaan menggunakan 6 nilai K yang berbeda, yaitu 50, 20, 10, 5, 3, dan 1. Proses pengkodean target label dapat dilihat pada Tabel *5* di bawah ini.

Tabel 5. Eksperimen Algoritma KNN

No.	Target Label Prediksi	Nilai K	Kode
1.	Halte	K=50	KNN K50 : H
2.	Halte	K=20	KNN K20 : H
3.	Halte	K=10	KNN K10 : H
4.	Halte	K=5	KNN K5 : H
5.	Halte	K=3	KNN K3 : H
6.	Halte	K=1	KNN K1 : H
7.	Sub-Koridor	K=50	KNN K50 : S-K
8.	Sub-Koridor	K=20	KNN K20 : S-K
9.	Sub-Koridor	K=10	KNN K10 : S-K
10.	Sub-Koridor	K=5	KNN K5 : S-K
11.	Sub-Koridor	K=3	KNN K3 : S-K
12.	Sub-Koridor	K=1	KNN K1 : S-K
13.	Koridor	K=50	KNN K50 : K
14.	Koridor	K=20	KNN K20 : K
15.	Koridor	K=10	KNN K10 : K
16.	Koridor	K=5	KNN K5 : K
17.	Koridor	K=3	KNN K3 : K
18.	Koridor	K=1	KNN K1 : K

2) Algoritma Decision tree

Pada kasus ini, *decision tree* memperluas sampai semua *leaves* memiliki nilai *pure* atau sampai semua *leaves* mengandung kurang dari min_sample_split yaitu 2. Sehingga tidak akan dilakukan batasan kedalaman *tree*. Kode eksperimen dengan menggunakan algoritma *Decision Tree* ditunjukkan pada Tabel 6.

Tabel 6. Kode Eksperimen Algoritma *Decision tree*

Label Prediksi	Atribut Waktu	
	Jam	Zona Waktu
Halte	DT J:H	DT ZW:H
Sub-Koridor	DT J:S-K	DT ZW:S-K
Koridor	DT J:K	DT ZW:K

3) Analisis Hasil Eksperimen

Dalam melakukan analisis hasil eksperimen dibagi menjadi 3 bagian yaitu analisis hasil eksperimen algoritma KNN, analisis hasil algoritma *decision tree,* dan analisis perbandingan algoritma KNN dan *decision tree*.

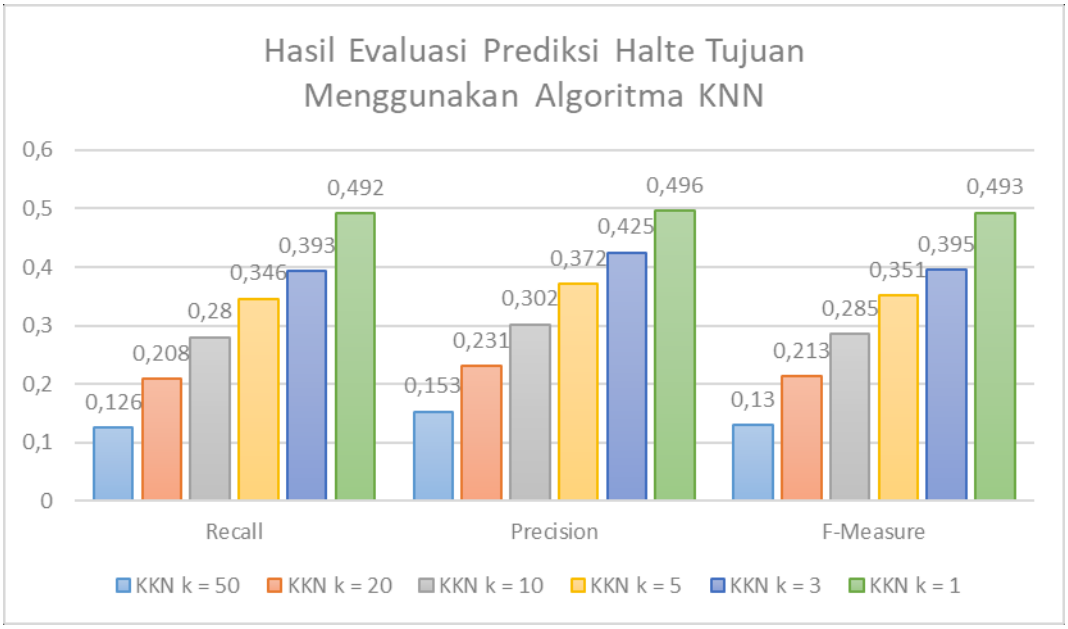
a) Analisis Hasil Eksperimen Algoritma KNN

Nilai K sangat mempengaruhi hasil prediksi tujuan. Nilai K=1 memberikan hasil yang signifikan lebih baik dibanding dengan K=3, K=10, K=20, dan K=50. Nilai K yang besar menghasilkan nilai *recall*, *precision*, dan f-measure yang lebih rendah

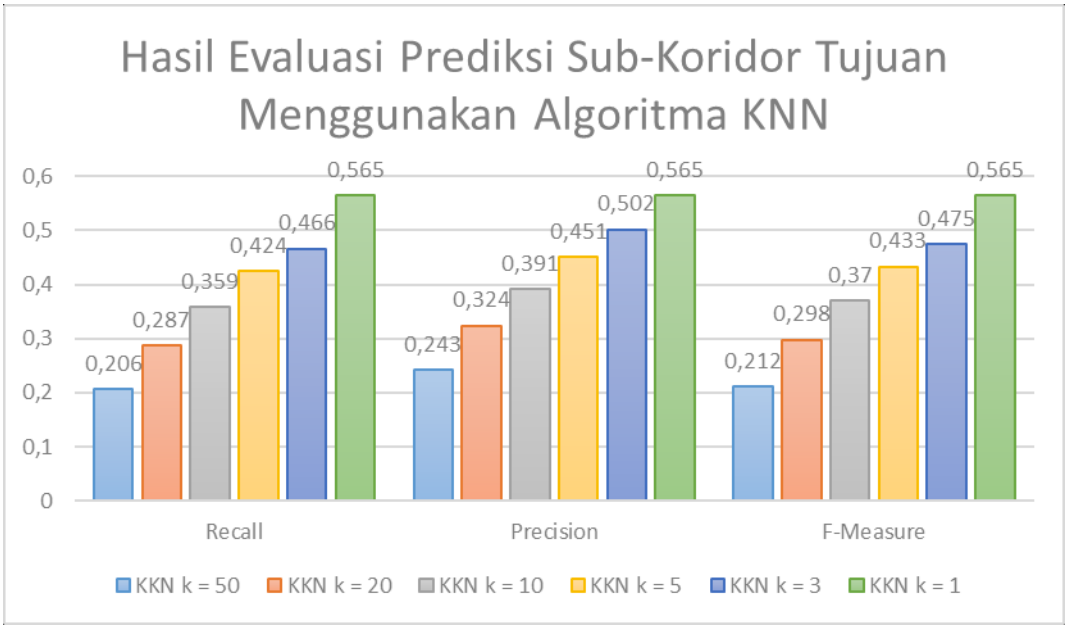
©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019

dibandingkan dengan nilai K yang kecil. Hal ini karena menggunakan nilai K yang besar memberikan bias dengan mengambil *data point* perjalanan individu penumpang yang berbeda. Sedangkan nilai K yang kecil memungkinkan untuk mengambil *data point* dengan perjalanan individu penumpang yang sama pada zona waktu yang sama.

Ketika K kecil, akan menahan wilayah prediksi yang diberikan dan memaksa *classifier* tidak melihat distribusi keseluruhan. Nilai kecil untuk K memberikan kecocokan paling fleksibel, yang akan memiliki biasa rendah tetapi varians tinggi. Disisi lain, K yang lebih tinggi melihat lebih banyak *data point* yang dijadikan pertimbangan di setiap prediksi dan karenanya lebih tahan terhadap pencilan. Nilai K yang lebih besar akan memiliki batas keputusan yang lebih halus yang berarti varians yang lebih rendah tetapi bias meningkat. Hasil evaluasi prediksi halte tujuan perjalanan penumpang dengan menggunakan algoritma KNN ditunjukkan pada Gambar 2, Gambar 3, dan Gambar 4.

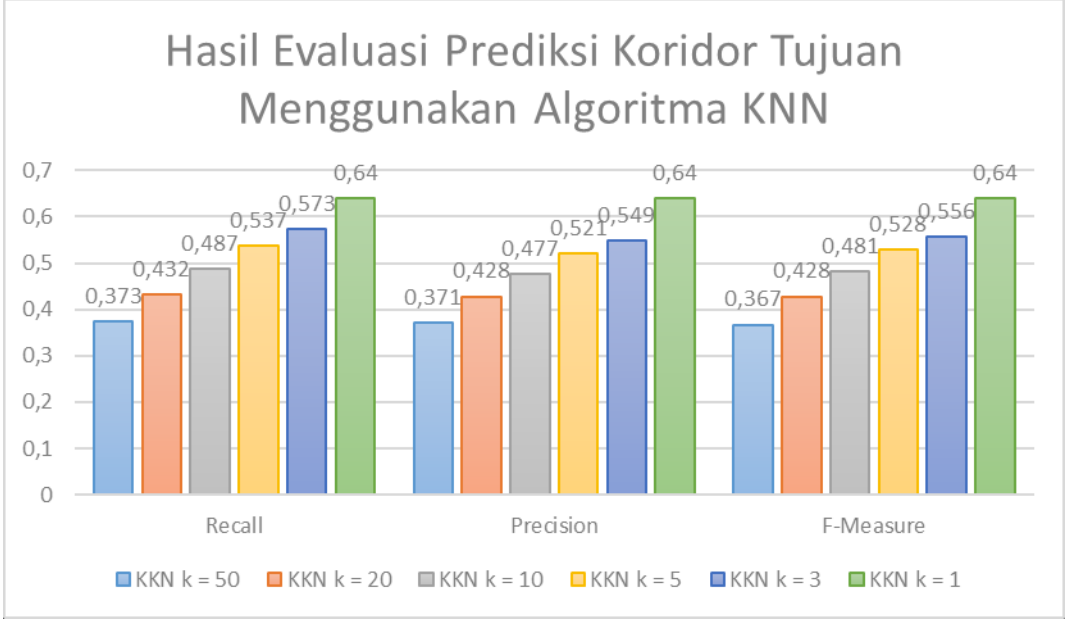


Gambar 2. Hasil Evaluasi Model Prediksi Halte Tujuan Menggunakan Algoritma KNN



©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019

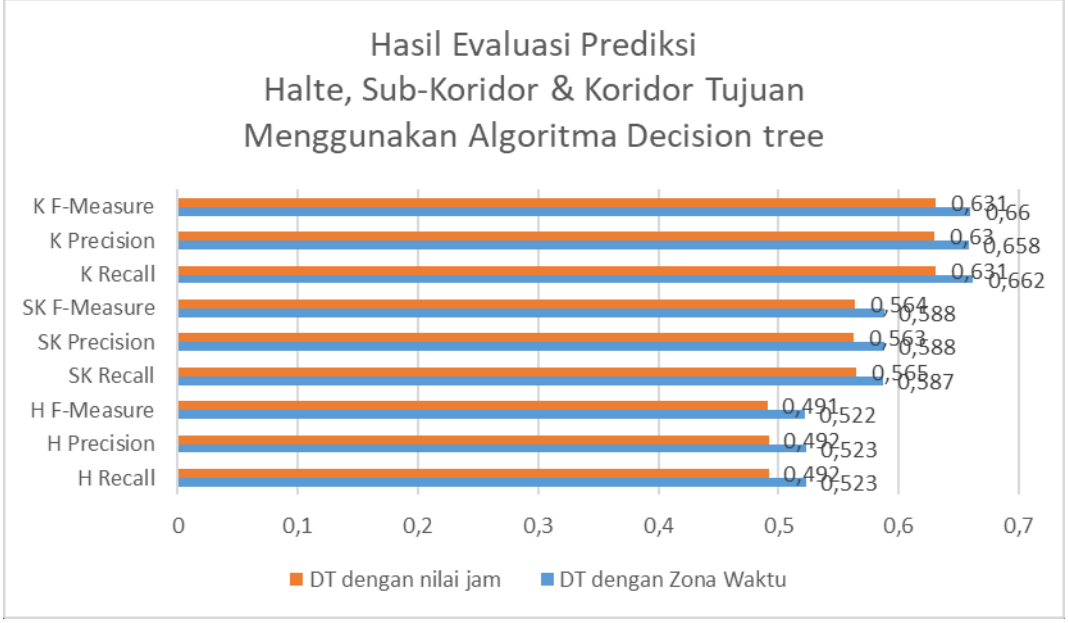
Gambar 3. Hasil Evaluasi Model Prediksi Sub-Koridor Tujuan Menggunakan Algoritma KNN



Gambar 4. Hasil Evaluasi Model Prediksi Koridor Tujuan Menggunakan Algoritma KNN

b) Analisis Hasil Eksperiment Algoritma Decision tree

Hasil eksperiment dengan menggunakan algoritma *decision tree* diketahui dengan menggunakan kategorisasi waktu menjadi zona waktu dapat memberikan nilai akurasi yang lebih baik dibandingkan dengan menggunakan nilai jam. Zona waktu dapat menaikkan nilai *recall, precision,* dan *f-measure* rata-rata 2,8%. Hasil evaluasi prediksi halte tujuan perjalanan penumpang dengan menggunakan algoritma *Decision Tree* ditunjukkan pada Gambar 5.



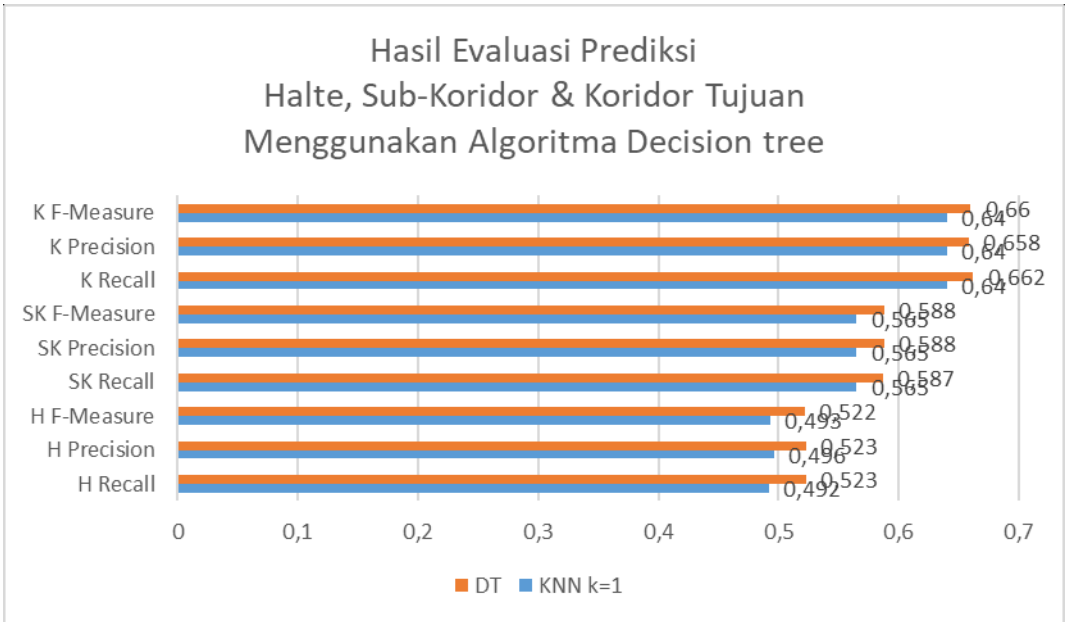
Gambar 5. Hasil Evaluasi Model Prediksi Tujuan Menggunakan Algoritma *Decision tree*

c) Analisis Perbandingan Hasil KNN Dan Decision tree

Rata-rata nilai *recall, precision,* dan *f-measure decision tree* 2,4% lebih tinggi dari nilai hasil evaluasi dengan KNN. Hal ini dikarenakan *decision tree* membagi setiap node

©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019

berdasarkan information gain tertinggi yang merupakan informasi perjalanan individu penumpang melalui atribut id_kartu. Sedangkan algoritma KNN merupakan "lazy learner" yang tidak melihat informasi gain dan melakukan prediksi dengan melihat kemiripan data point yang dalam hal ini melihat kemiripan semua atribut secara merata walaupun dengan menggunakan nilai K terendah. Hasil perbandingan evaluasi model prediksi halte tujuan perjalanan penumpang dengan menggunakan algoritma KNN dan *Decision Tree* ditunjukkan pada Gambar 6.



Gambar 6. Perbandingan Hasil Evaluasi Model Prediksi Tujuan Menggunakan Algoritma KNN dan *Decision tree*

4) Evaluasi model

Pengujian model terhadap model prediksi dilakukan dengan menggunakan *congfusion matrix* untuk menghitung *recall, precision,* dan *f-measure*.

a) KNN

Hasil evaluasi model yang didapatkan dengan menggunakan algoritma KNN dapat dilihat pada Tabel 7.

Tabel 7. Hasil Evaluasi KNN

No.	Kode Skenario	*Recall*	*Precision*	*F-measure*
1.	KNN K50 : H	0.126	0.153	0.129
2.	KNN K20 : H	0.208	0.231	0.213
3.	KNN K10 : H	0.280	0.302	0.285
4.	KNN K5 : H	0.346	0.372	0.351
5.	KNN K3 : H	0.393	0.425	0.395
6.	KNN K1 : H	**0.492**	**0.496**	**0.493**
7.	KNN K50 : S-K	0.206	0.243	0.211
8.	KNN K20 : S-K	0.287	0.324	0.298
9.	KNN K10 : S-K	0.359	0.391	0.370
10.	KNN K5 : S-K	0.424	0.451	0.433
11.	KNN K3 : S-K	0.466	0.502	0.475
12.	KNN K1 : S-K	**0.565**	**0.565**	**0.565**
13.	KNN K50 : K	0.373	0.370	0.366
14.	KNN K20 : K	0.431	0.427	0.427
15.	KNN K10 : K	0.487	0.477	0.481
16.	KNN K5 : K	0.537	0.537	0.537

©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019

b) Decision tree

Hasil evaluasi yang didapatkan pada tiap eksperimen, sesuai dengan skenario yang telah didefinisikan pada Tabel *6* dapat dilihat pada Tabel *8* berikut ini.

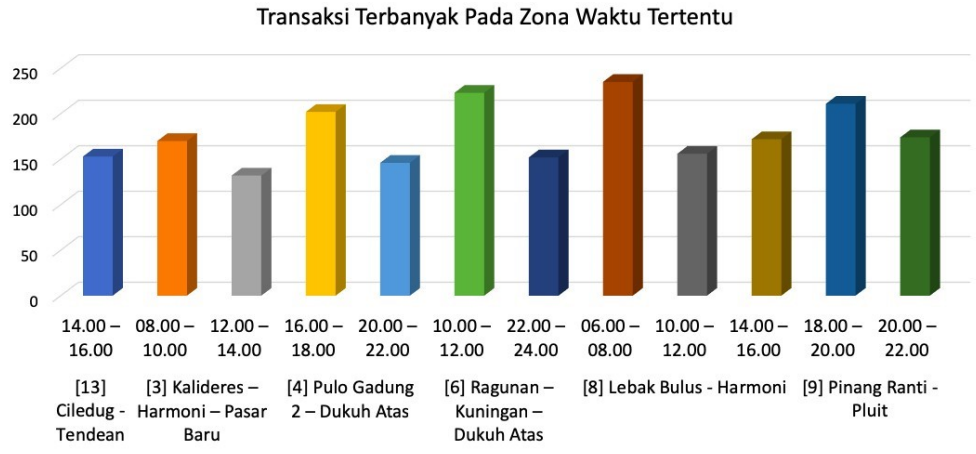
Tabel 8. Evaluasi Kinerja dengan *decision tree*

No.	Kode Skenario	*Recall*	*Precision*	*F-measure*
1.	DT J:H	0.492	0.492	0.492
2.	DT ZW:H	**0.523**	**0.523**	**0.522**
3.	DT J:S-K	0.565	0.563	0.564
4.	DT ZW:S-K	**0.587**	**0.588**	**0.588**
5.	DT J:K	0.631	0.630	0.631
6.	DT ZW:K	**0.662**	**0.658**	**0.660**

5) Informasi Strategis dari Analisis Menggunakan Decision tree

Hasil analisis menggunakan *decision tree* dapat dilihat bahwa terdapat beberapa pasangan koridor dan sub-koridor yang menjadi asal-tujuan yang paling ramai pada jam-jam tertentu. Sehingga dapat dilakukan beberapa analisis penyebab ramainya koridor tersebut. Kemudian dari hasil analisis tersebut, dapat dilakukan beberapa keputusan strategis yang berkaitan dengan infrastruktur maupun pelayanan.

Dari hasil evaluasi dengan *decision tree*, Tabelmenunjukkan bahwa evaluasi kinerja model pembelajaran sudah cukup baik, sehingga dapat dijadikan landasan untuk pengambilan keputusan.. menunjukkan beberapa pasangan Koridor yang memiliki banyak transaksi *tap in-out* pada zona-zona waktu tertentu. Dari hasil analisis berikut ini, dapat dijadikan sebagai rekomendasi atau pendukung penyusunan analisis lanjut untuk merumuskan strategi. Misalnya dapat dilakukan dengan survei langsung, untuk mengobservasi keadaan langsung di lapangan. Sehingga dapat dilihat *gap* antara hasil analisis dengan hasil survei



Gambar 7. Sebaran Total Transaksi Terbanyak

Pada Gambar 7 dapat dilihat jam-jam sibuk transaksi *tap in-out* terbanyak. Pasangan koridor Lebak Bulus – Harmoni pada zona waktu 06.00 – 08.00 merupakan koridor dengan jumlah transaksi *tap in-out* paling banyak.

©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019

IV. KESIMPULAN

- Berdasarkan nilai *information gain*, mengkonversi waktu menjadi kategorikal efektif untuk meningkatkan kinerja model prediksi. Selain itu informasi perjalanan individu penumpang melalui id kartu merupakan atribut yang paling berkorelasi untuk melakukan prediksi tujuan.
- Dibandingkan dengan KNN, kinerja *predictive model* dengan *decision tree* memberikan hasil kinerja yang lebih baik dilihat dari nilai *precision*, *recall*, dan f- *measure*. Hal ini dikarenakan *decision tree* membagi *node* berdasarkan *information gain* tertinggi yaitu informasi perjalanan individu penumpang yang paling berkorelasi dengan prediksi tujuan penumpang.

REFERENCES

[1] T. Wen, C. Cai, L. Gardner, S. T. Waller, V. Dixit, and F. Chen, "Estimation of sparse O–D matrix accounting for demand volatility," *IET Intell. Transp. Syst.*, 2018.

[2] C. Sun, L. Cheng, X. Luan, Q. Tu, and J. Ma, "Subnetwork origin-destination matrix estimation considering network topology," *Dongnan Daxue Xuebao (Ziran Kexue Ban)/Journal Southeast Univ. (Natural Sci. Ed.*, 2017.

[3] M. Bagchi and P. R. White, "What role for smart-card data from bus systems?," *Proc.Inst. Civ. Eng. - Munic. Eng.*, 2009.

[4] W. Wang, J. Attanucci, and N. Wilson, "Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems," *J. Public Transp.*, 2015.

[5] S. Tao, J. Corcoran, I. Mateo-Babiano, and D. Rohde, "Exploring Bus Rapid Transit passenger travel behaviour using big data," *Appl. Geogr.*, 2014.

[6] T. Li, D. Sun, P. Jing, and K. Yang, "Smart card data mining of public transport destination: A literature review," *Information (Switzerland)*. 2018.

[7] O. J. Ibarra-Rojas, F. Delgado, R. Giesen, and J. C. Muñoz, "Planning, operation, and control of bus transport systems: A literature review," *Transportation Research Part B: Methodological*. 2015.

[8] Widyawan, B. Prakasa, D. W. Putra, S. S. Kusumawardani, B. T. Y. Widhiyanto, and F. Habibie, "Big data analytic for estimation of origin-destination matrix in Bus Rapid Transit system," in *Proceeding - 2017 3rd International Conference on Science and Technology-Computer, ICST 2017*, 2017.

[9] J. J. Barry, R. Newhouser, A. Rahbee, and S. Sayeda, "Origin and Destination Estimation in New York City with Automated Fare System Data," *Transp. Res. Rec. J. Transp. Res. Board*, 2007.

[10] J. Zhao, A. Rahbee, and N. H. M. Wilson, "Estimating a rail passenger trip origin-destination matrix using automatic data collection systems," *Comput. Civ. Infrastruct. Eng.*, 2007.

[11] C. Manasseh and R. Sengupta, "Predicting driver destination using machine learning techniques," in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2013.

[12] L. H. Tran, M. Catasta, L. K. McDowell, and K. Aberer, "Next Place Prediction using Mobile Data," *Proc. Mob. Data Chall. Work. (MDC 2012)*, 2012.

[13] A. Lo Duca, C. Bacciu, and A. Marchetti, "A K-nearest neighbor classifier for ship route prediction," in *OCEANS 2017 - Aberdeen*, 2017.

[14] V. Roşca, E. Onica, P. Diac, and C. Amariei, "Predicting Destinations by Nearest Neighbor Search on Training Vessel Routes," 2018.

[15] U. Al-Faruqi, "Survey Paper : Future Service in Industry 5 . 0," vol. 02, no. 01, pp. 67–79, 2019.

©Asosiasi Prakarsa Indonesia Cerdas (APIC) - 2019