



MODEL ESTIMASI BANGKITAN PERGERAKAN MODA LAUT MENGGUNAKAN METODE REGRESI LINIER DAN *RANDOM FOREST*

Fadhlullah Apriandy^{a,*}, Sugiarto Sugiarto^a, Sofyan M Saleh^a, Lulusi Lulusi^a

^aJurusan Teknik Sipil, Fakultas Teknik, Universitas Syiah Kuala, Banda Aceh

*Corresponding author, email address: fadhlullah.apriandy@unsyiah.ac.id

ARTICLE INFO	ABSTRACT
<i>Article History:</i> Recieved 7 November 2021 Accepted 28 December 2021 Online 30 December 2021 <i>Keywords:</i> Trip generation Trip attraction Sea transportation Linear regression Random forest Socio-demographics	The need for people and goods movement across regions have been increasing substantially over the years. Sea transportations plays pivotal role in allowing this trans-national movement. The ferry line connecting Langsa, Indonesia and Penang, Malaysia had been established to help people and good across borders, albeit no longer operating. This study aims to estimate trip generation and attraction between these ports and their surrounding areas, and suggest whether this ferry line should reoperate. In doing so, random forest was utilised alongside linear regression model. Random forest has been increasingly popular among researchers in making estimates or preditions, whether calculating numerical or categorical data. To develop both models, historical trip and socio-economics data were employed. It was found that Gross Domestic Regional Product (GDRP) and population significantly affect trip generation and attraction. Both variables shows positive effect on trip generation and attraction: an increase in either should increase trip generation and attraction. Similar to linear regression model, random forest model show an excellent predictive capability. With a sophisticated model incorporating more variables, it is expected that random forest should well exceed linear regression predicitive capability. As population and GDRP are expected to continue to grow, it is suggested that stakeholders should reoperate this Langsa-Penang ferry line.

©2021 Magister Teknik Sipil Unsyiah. All rights reserved

1. PENDAHULUAN

Seiring dengan semakin terjangkaunya moda transportasi udara, pergerakan orang dan barang antar-regional semakin berkembang pesat. Ditambah lagi dengan penerapan Kawasan Perdagangan Bebas Tionkok-Asean (CAFTA), sekat pemisah antar negara semakin pudar. Meskipun demikian, ada potensi besar akan transporatsi laut, terutama bagi pergerakan barang, karena wilayah ini memiliki sumber daya laut yang melimpah yang dapat dimanfaatkan bagi sarana transportasi laut.

Lembaga Perserikatan Bangsa-Bangsa (PBB) yang fokus kepada negara berkembang di dunia perdagangan (United Nations Conference on Trade and Development/UNCTAD) mengatakan bahwa sekitar 90% dari perdagangan global menggunakan moda transportasi laut. Analisis Kepadatan Pergerakan Kapal Laut secara global oleh Tournadre (2014) menunjukkan bahwa lalu lintas transportasi laut di tahun 2014 meningkat menjadi empat kali lipat dari level di awal dekade 90-an, dengan Samudra India dan Pasifik Barat mengalami peningkatan paling pesat. Peningkatan kepadatan ini erat kaitannya dengan karakteristik dari moda transportasi laut itu sendiri, seperti penurunan biaya transportasi laut, berkurangnya waktu pengiriman, dan meningkatnya keamanan pengiriman (Cordón-Lagares dan García-Ordaz, 2020). Selain itu, transportasi laut menghasilkan emisi CO₂ per ton berat kargo yang lebih rendah dibandingkan

moda transportasi lainnya, sehingga lebih ramah lingkungan (Grant dan Elliot, 2018). Volume kargo yang lebih besar, baik dari sisi ukuran maupun berat, juga memainkan peranan penting dalam pertimbangan orang untuk memilih moda transportasi laut.

Transportasi pada umumnya rentan terhadap perubahan kebijakan dan perkembangan teknologi. Akibatnya, perlu adanya studi akan prediksi untuk memperkirakan kebutuhan transportasi di masa depan (*transport demand*) seiring dengan pembaruan kebijakan dan perkembangan teknologi yang pesat. Estimasi tersebut dapat digunakan untuk menentukan dan menganalisis apakah diperlukan beberapa penyesuaian, baik dari sisi teknis maupun manajemen. Beberapa faktor yang banyak digunakan dalam mengestimasi kebutuhan transportasi antara lain Produk Domestik Bruto (PDB), jumlah penduduk, pengeluaran rumah tangga, harga bahan bakar, tarif tiket, dan lain-lain.

Penelitian ini difokuskan pada rute laut internasional yang menghubungkan Langsa, Indonesia dengan Penang, Malaysia. Pertama kali dibuka pada tahun 2013, bertujuan untuk meningkatkan dan memperkuat kerja sama bilateral di bidang politik, ekonomi, sosial, budaya, investasi, dan pariwisata. Akan tetapi, rute ini berhenti beroperasi setelah hanya beberapa bulan. Oleh karenanya, hasil dari studi ini diharapkan dapat menjadi pertimbangan bagi para *stakeholders* apakah rute ini layak untuk dioperasikan kembali.

Selain menggunakan metode yang sudah umum yaitu regresi linier, studi ini juga menggunakan metode pembelajaran mesin (*machine learning*) yaitu *random forest* dalam mengestimasi kebutuhan pergerakan. *Random forest* semakin digemari oleh para peneliti dalam melakukan prediksi, baik prediksi tersebut berupa data numerik maupun data kategorikal. Kelebihan dari *random forest* adalah kemampuan prediksinya yang mampu menghasilkan tingkat akurasi yang lebih baik dibanding metode konvensional seperti regresi linier.

2. KAJIAN PUSTAKA

Beberapa faktor yang banyak digunakan dalam mengestimasi kebutuhan transportasi antara lain Produk Domestik Bruto (PDB), jumlah penduduk, pengeluaran rumah tangga, harga bahan bakar, tarif tiket, dan lain-lain. Variabel-variabel ini dianggap dominan dan memainkan peranan penting dalam mencerminkan kebutuhan transportasi di masa depan. Selain itu, Tsui dan Fung (2016) mengatakan bahwa terdapat korelasi yang kuat antara PDB dan kebutuhan transportasi, oleh karenanya dianggap sebagai variabel yang paling cocok dalam mencerminkan kebutuhan transportasi.

Dalam studi ini, fitur yang digunakan untuk memperkirakan bangkitan dan tarikan pergerakan di masa mendatang adalah jumlah penduduk dan Produk Domestik Regional Bruto (PDRB). Variabel-variabel ini dianggap eksogen. PDRB bervariasi dari waktu ke waktu karena fluktuasi harga barang dan jasa serta perubahan volume. Dataset ini dikumpulkan dari Badan Pusat Statistik (BPS Langsa) Langsa (BPS Langsa, 2020). Data sebelum tahun 2010 tidak dipertimbangkan karena sesuai dengan rekomendasi PBB, perhitungan tahun dasar produk domestik bruto telah diubah dari tahun 2000 menjadi 2010 dengan mempertimbangkan keadaan luar biasa seperti CAFTA, sistem pencatatan perdagangan internasional, dan perluasan layanan pasar modal. Berdasarkan data yang digunakan, studi ini dianggap sebagai metode berbasis ekonomi dan deret waktu (*time series*).

Pemodelan kebutuhan transportasi pada umumnya menggunakan pendekatan statistik seperti regresi, model gravitasi, dan metode *time-series* (Profillidis dan Botzoris, 2019). Pendekatan-pendekatan ini, yang perlu dibuktikan secara matematis, mewajibkan data untuk memenuhi asumsi yang sudah melekat sebelumnya. Tidak seperti metode konvensional, teknik pembelajaran mesin membebaskan data dari mengikuti asumsi dasar seperti distribusi variabel (Aghaabbasi dkk, 2020) dan memberi kebebasan bagi algoritma komputer untuk belajar dari observasi dan mengamati pola. Hal ini menawarkan fleksibilitas bagi para peneliti dalam mengembangkan model prediksi (Zhao dkk, 2020).

Metode berbasis pembelajaran mesin dapat digunakan di antaranya untuk tujuan deskriptif dengan menggambarkan status sistem saat ini; tujuan prediktif dengan memprediksi kondisi sistem berikutnya; atau tujuan preskriptif dengan mengusulkan perbaikan sistem (Barua dkk, 2020). Algoritma pembelajaran mesin

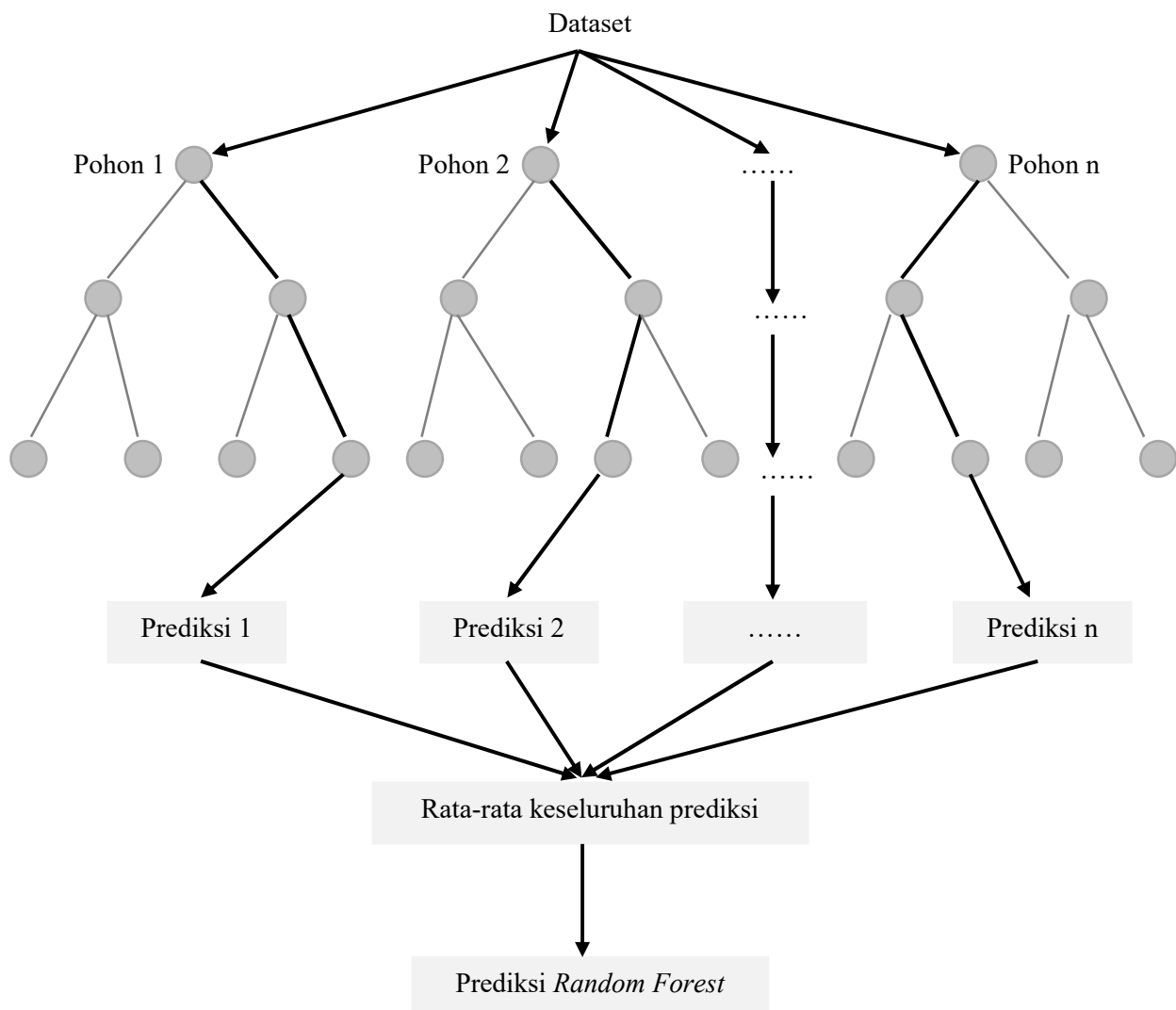
mengajarkan komputer untuk mereplikasi cara paling mendasar dari bagaimana manusia belajar, yaitu belajar dari pengalaman. Beberapa penelitian terdahulu seperti yang dilakukan oleh Barua dkk (2020) dan Alekseev dan Seixas (2009) telah menunjukkan bahwa pendekatan pembelajaran mesin sering mengungguli teknik pemodelan konvensional. Beberapa metode pembelajaran mesin yang paling umum digunakan dalam banyak penelitian antara lain K-means Clustering, *Artificial Neural Network* (ANN), dan *Random Forest* (RF). Meningkatnya minat dalam penggunaan metode pembelajaran mesin di berbagai bidang penelitian dikaitkan dengan kemampuannya untuk memecahkan permasalahan dengan beragam tipe data (Cheng dkk, 2020). Studi terdahulu menunjukkan metode pembelajaran mesin lebih unggul dibandingkan metode konvensional dalam kemampuan prediksi (Hagenauer dan Helbich, 2017; Zhao dkk, 2020; Cheng dkk, 2020; Yan dkk, 2020).

Dalam memodelkan hubungan antara variabel bebas dan terikat, metode regresi linier menggunakan rumus 1 sebagai berikut.

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + \varepsilon$$

Di mana:

- y = jumlah bangkitan/tarikan perjalanan
- a_i = koefisien dari variabel bebas
- x_i = variabel bebas meliputi PDRB, populasi
- ε = galat (*error term*)



Gambar 1. Sistem pengambilan keputusan *Random Forest*

Metode *Random Forest* terinspirasi dari suatu kombinasi pepohonan yang membentuk suatu sistem hutan. Cara *Random Forest* bekerja dalam melihat pola data adalah dengan meminta masukan dari setiap individu pohon dalam membuat suatu prediksi, di mana galat (*error*) dari prediksi akan mengerucut dengan meningkatnya jumlah pepohonan dalam sistem itu sendiri (Breiman, 2001). Dengan kata lain, suatu sistem *Random Forest* yang terdiri dari pepohonan dengan jumlah yang banyak diharapkan mampu menghasilkan suatu prediksi yang lebih akurat. Akan tetapi, kemampuan *Random Forest* dalam meminimalisir *error* dalam membuat generalisasi data sangat dipengaruhi oleh kekuatan dari setiap pohon dan korelasi di antara mereka. *Random Forest* membagi dataset menjadi beberapa sub-sampel secara acak (*random*) dan menggunakan rata-rata untuk meningkatkan kemampuan prediksi. Skema pengambilan keputusan *Random Forest* dapat dilihat pada Gambar 1.

3. METODE PENELITIAN

Penelitian ini menggunakan metode regresi linier dan *Random Forest* untuk memodelkan bangkitan perjalanan dan tarikan perjalanan dengan menggunakan PDRB dan populasi sebagai variabel bebas. Kedua metode dibangun menggunakan algoritma komputer berbasis Python yang dijalankan di *Scientific Python Development Environment (Spyder)*. Modul *dmatrixes* dari *patsy* digunakan pada tahap pengolahan data. *Dmatrixes* berguna dalam memisahkan variabel independen dan dependen dan menambahkan nilai konstan dalam analisis regresi. *Statsmodels* digunakan untuk membantu menjalankan metode regresi linier. *Statsmodels* memiliki fungsi *Ordinary Least Square* (OLS) yang dapat memodelkan hubungan linier antara variabel independen dan dependen. Eksekusi dari algoritma tersebut akan menghasilkan intersep (konstanta), koefisien setiap variabel independen, dan parameter statistik penting seperti *R-squared* dan *p-value*. Uji tingkat kepercayaan 5% digunakan dalam penelitian ini. Sedangkan metode *Random Forest* dapat dijalankan menggunakan modul dari *sci-kit learn*. Parameter yang berperan penting dalam sistem *Random Forest* di antaranya adalah jumlah individu pohon (*estimator*).

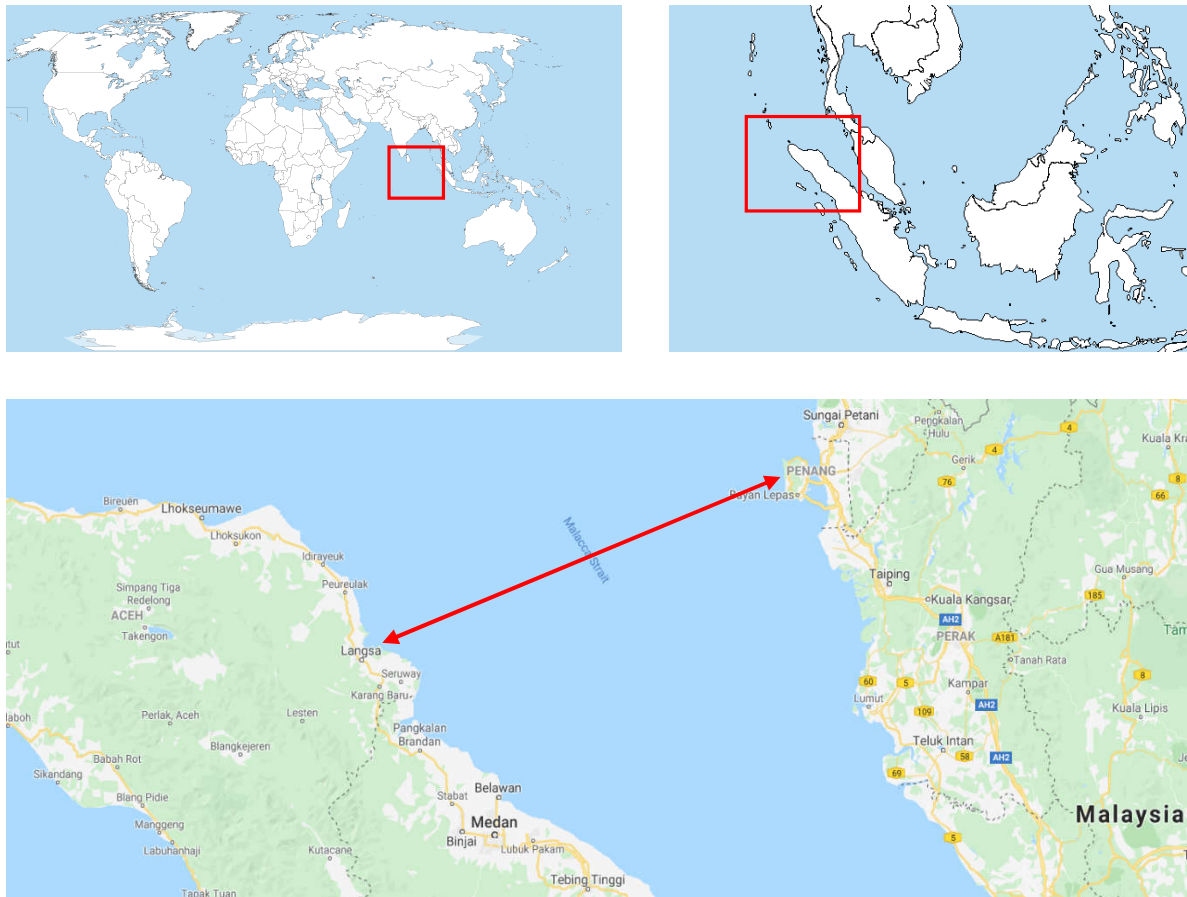
Penelitian ini difokuskan pada rute pelayaran internasional Langsa-Penang seperti digambarkan pada Gambar 2. Perjalanan dimulai dari Pelabuhan Kuala Langsa, Langsa, Aceh, Indonesia, kemudian berlabuh di Pelabuhan Penang, Penang, Malaysia, dan sebaliknya. Pertama kali dibuka pada tahun 2013, pelayaran ini bertujuan untuk meningkatkan dan memperkuat kerja sama bilateral di bidang politik, ekonomi, sosial, budaya, investasi, dan pariwisata. Ini adalah satu-satunya moda transportasi langsung yang menghubungkan Langsa dan Penang, dan satu-satunya kapal penumpang berjadwal yang menghubungkan bagian utara Pulau Sumatera dengan Pulau Penang. Kapal Feri Kenangan-3 berkapasitas 192 penumpang yang dimiliki oleh Fast Ferry Ventures Sdn Bhd, adalah satu-satunya kapal yang melayani rute tersebut. Kapal beroperasi dua kali seminggu, berangkat dari Penang pukul 8.30 pagi pada hari Senin dan Rabu, serta meninggalkan Langsa setiap Selasa dan Kamis pukul 09.00. Dengan jarak tempuh sekitar 300 km, perjalanan satu arah biasanya memakan waktu 6 hingga 7 jam, tergantung kondisi laut dan cuaca. Biaya perjalanan pulang-pergi adalah Rp. 500.000. Akan tetapi, setelah beroperasi selama beberapa bulan, rute pelayaran ini berhenti beroperasi. Hal yang menjadi pertimbangan di antaranya adalah rendahnya okupansi kapal.

4. HASIL DAN PEMBAHASAN

Rangkuman model regresi linier ditunjukkan pada Tabel 1 dan Tabel 2. Sekilas, kecilnya nilai koefisien masing-masing variabel independen terkesan menunjukkan ketidaksigifikannya. Namun, *p-value* yang rendah menunjukkan bahwa variabel independen, dalam hal ini PDRB dan populasi, memiliki hubungan yang kuat dengan bangkitan perjalanan dan daya tarik perjalanan. Koefisien yang kecil bukan bermakna tidak berarti. Ini mungkin hasil dari data yang tidak seimbang di mana populasi adalah data angka dengan enam digit, PDRB adalah angka tujuh digit, sedangkan perjalanan itu sendiri adalah angka lima digit. Hasil regresi menunjukkan bahwa, untuk setiap 10.000 peningkatan populasi, diperkirakan akan terjadi 121 peningkatan bangkitan perjalanan dan 33 tarikan perjalanan. Di sisi lain, pertumbuhan PDRB sebesar satu juta rupiah diperkirakan akan menghasilkan 2700 bangkitan pergerakan dan 700 tarikan pergerakan.

Dengan potensi pertumbuhan penduduk dan PDRB, studi ini menemukan bahwa rute pelayaran internasional Langsa-Penang layak untuk dioperasikan kembali.

Metode *Random Forest* menghasilkan nilai *R-squared* yang nyaris sempurna baik pada model bangkitan pergerakan maupun tarikan pergerakan. Metode regresi linier juga menghasilkan nilai *R-squared* serupa. Hal ini terjadi dikarenakan jumlah variabel bebas yang digunakan sedikit, yaitu hanya dua. Jika variabel bebas yang digunakan lebih banyak dan beragam, diperkirakan bahwa model *Random Forest* akan menghasilkan prediksi yang lebih akurat dibandingkan model konvensional, seperti yang ditunjukkan oleh studi terdahulu. Akan tetapi, tidak ada luaran koefisien dari metode *Random Forest*. Hal ini membatasi pemahaman yang baik akan besarnya pengaruh dari setiap variabel. Inilah salah satu kekurangan dari metode pembelajaran mesin, meskipun di sisi lain metode ini memiliki kemampuan akurasi yang sangat baik. Pembelajaran mesin terkesan bekerja layaknya sebuah kotak hitam (*black box*).



Gambar 2. Lokasi penelitian

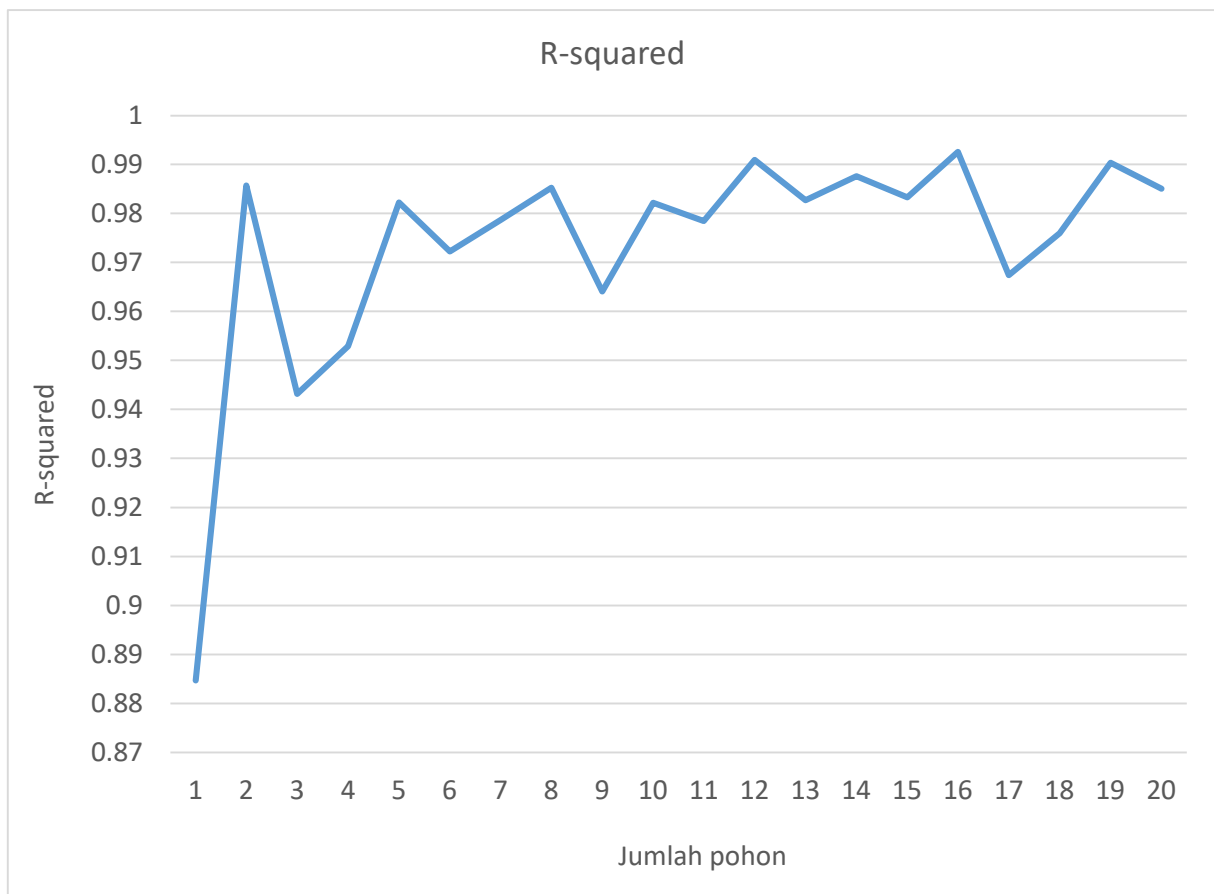
Table 1. Model bangkitan pergerakan

Model	Variabel bebas	Koefisien	<i>P-value</i>	<i>R-squared</i>
Regresi linier	Intersep	-1307.88	0.038	0.99
	Populasi	0.0121	0.043	
	PDRB	0.0027	0.000	
<i>Random Forest</i>	-	-	-	0.99

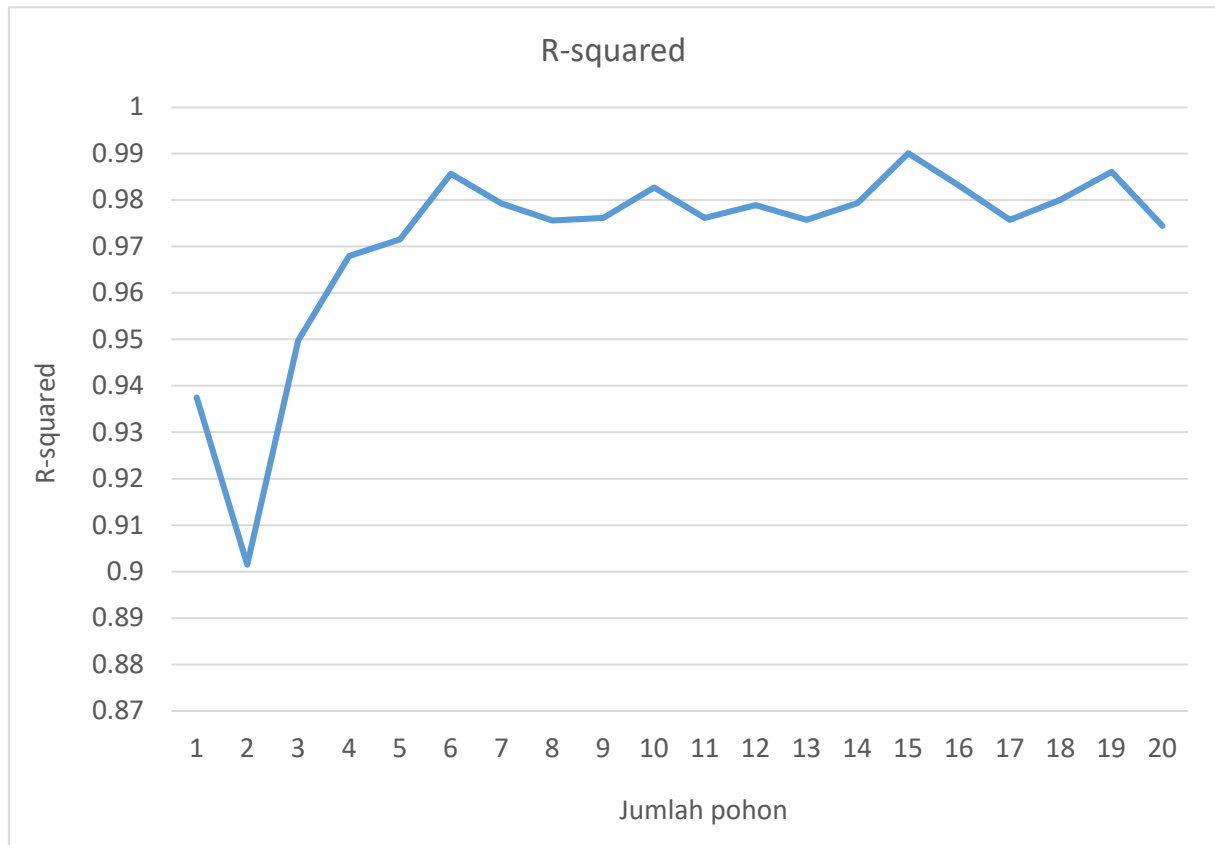
Table 2. Model tarikan pergerakan

Model	Variabel bebas	Koefisien	<i>P-value</i>	<i>R-squared</i>
Regresi linier	Intersep	-358.385	0.038	0.99
	Populasi	0.0033	0.043	
	PDRB	0.0007	0.000	
<i>Random Forest</i>	-	-	-	0.99

Gambar 3 dan 4 menunjukkan nilai *R-squared* yang diperoleh melalui metode *Random Forest* pada model bangkitan dan tarikan pergerakan. Dilakukan beberapa percobaan jumlah pohon. Terlihat bahwa nilai *R-squared* yang dihasilkan fluktuatif. Nilai *R-squared* yang diperoleh pada mulanya rendah dengan jumlah pohon yang sedikit (<5), kemudian meningkat seiring dengan bertambahnya jumlah pohon, meskipun terjadi fluktuasi nilai *R-squared*. Akan tetapi, setelah beberapa percobaan, dapat dilihat bahwa nilai *R-squared* mulai stabil di level nyaris sempurna dengan pohon berjumlah belasan. Fluktuasi ini terjadi karena model *Random Forest*, sesuai dengan namanya, bekerja dengan membagi dataset mejadi sub-sampel secara acak (*random*).



Gambar 3. *R-squared* model bangkitan pergerakan



Gambar 4. *R-squared* model tarikan pergerakan

Nilai *R-squared* yang diperoleh mengindikasikan bahwa baik metode regresi linier maupun *Random Forest* mampu memodelkan data dengan nyaris sempurna. Dari perspektif statistik, hal ini dianggap sebagai model yang ideal. Namun, ini juga mengindikasikan kemungkinan terjadinya *overfitting* yang sering terjadi pada riset dengan jumlah data yang minim.

5. KESIMPULAN

Studi ini menunjukkan bahwa penerapan metode pembelajaran mesin dalam pemodelan dan prediksi bangkitan dan tarikan pergerakan dapat dikatakan sebagai instrumen yang bermanfaat untuk digunakan. Hal ini ditandai dengan kemampuan prediksi yang sangat baik dari model *Random Forest*, meskipun di lain sisi, ketiadaan koefisien dari tiap variabel menjadikan terbatasnya pemahaman akan besarnya pengaruh masing-masing faktor terhadap bangkitan dan tarikan pergerakan. Studi ini menemukan bahwa data sosial ekonomi meliputi populasi dan PDRB berpengaruh terhadap jumlah bangkitan dan tarikan pergerakan. Dengan potensi pertumbuhan penduduk dan PDRB, studi ini menemukan bahwa rute pelayaran internasional Langsa-Penang layak untuk dioperasikan kembali. Nilai *R-squared* yang dihasilkan baik oleh model regresi linier maupun *Random Forest* sangatlah tinggi, yang mengindikasikan model yang ideal di satu sisi dan kemungkinan terjadinya *overfitting* akibat minimnya jumlah data di sisi yang lain.

DAFTAR PUSTAKA

- Aghaabbasi, M., Shekari, Z. A., Shah, M. Z., Olakunle, O., Armaghani, D. J., Moeinaddini, M. 2020. Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. *Transportation Research Part A: Policy and Practice*, 136, pp. 262-281.

- Alekseev, K. P. G., Seixas, J. M. 2009. A multivariate neural forecasting modeling for air transport–preprocessed by decomposition: a Brazilian application. *Journal of Air Transport Management*, 15(5), pp. 212-216.
- Barua, L., Zou, B., Zhou, Y. 2020. Machine learning for international freight transportation management: a comprehensive review. *Research in Transportation Business & Management*. 34, p. 100453.
- BPS Langsa. 2020. *Langsa Municipality in Figures 2020*. BPS Langsa, Langsa, Indonesia.
- Breiman, L. 2001. Random forests. *Machine learning*, 45(1), pp. 5-32.
- Cheng, L., Chen, X., De Vos, J., Lai, X., Witlox, F. 2019. Applying a random forest method approach to model travel mode choice behavior. *Travel Behaviour and Society*, 14, pp. 1-10.
- Cordón-Lagares, E., García-Ordaz, F. 2020. Factors affecting the survival of maritime goods transport firms in Spain. *Research in Transportation Business & Management*. 37, p. 100520.
- Grant, D. B., Elliott, M. 2018. A proposed interdisciplinary framework for the environmental management of water and air-borne emissions in maritime logistics. *Ocean & coastal management*, 163, pp.162-172.
- Hagenauer, J., Helbich, M. 2017. A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, 78, pp. 273-282.
- Profillidis, V. A., Botzoris, G. N. 2019. *Modeling of Transport Demand*. Elsevier, Oxford, UK.
- Tournadre, J. 2014. Anthropogenic pressure on the open ocean: The growth of ship traffic revealed by altimeter data analysis. *Geophysical Research Letters*, 41(22), pp.7924-7932.
- Tsui, W. H., Fung, M. K. 2016. Analysing passenger network changes: The case of Hong Kong. *Journal of Air Transport Management*, 50, pp. 1-11.
- Yan, X., Liu, X., Zhao, X. 2020. Using machine learning for direct demand modeling of ridesourcing services in Chicago. *Journal of Transport Geography*, 83, p.102661.
- Zhao, X., Yan, X., Yu, A., Van Hentenryck, P. 2020. Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel behaviour and society*. 20, pp. 22-35.