



PERBANDINGAN *XGB REGRESSOR* DENGAN ALGORITMA LAIN UNTUK PREDIKSI TARIF TOL

Said Al Khairi¹, Ahmad Rio Adriansyah², Lukman Rosyidi³

^{1,2,3} Teknik Informatika, Sekolah Tinggi Teknologi Terpadu Nurul Fikri

Jakarta Selatan, DKI Jakarta, Indonesia 12640

said20277ti@student.nurulfikri.ac.id, arasy@nurulfikri.ac.id, lukman@nurulfikri.ac.id,

Abstract

In recent years, toll roads in Indonesia have grown rapidly, many of which were built to facilitate traffic in developed areas and improve the distribution of goods and services to support economic growth. In addition, toll roads play an important role as part of efforts to improve connectivity between cities and regions and accelerate community mobility. Many benefits of toll roads have been felt by the people of Indonesia such as, the Jagorawi toll road which smooths traffic so as to shorten the travel time from one region to another, and many more. The purpose of this research is to create a machine learning prediction of toll road tariffs to provide a reference to the public, optimise toll tariffs in Indonesia, and provide input on toll tariffs as a consideration for the relevant government. This research approach is quantitative using linear regression with *XGB Regressor* algorithm. The results of making machine learning toll tariff predictions are quite accurate where the accuracy test results using the root mean squared error (RMSE) metric are at 3390.691, with the testing results showing that there are several predicted tariffs that match the original tariff.

Keywords: machine learning, fare prediction, linear regression, toll rates, *XGB Regressor*.

Abstrak

Beberapa tahun terakhir jalan tol di Indonesia telah berkembang pesat, banyak jalan tol di Indonesia dibangun guna memperlancar lalu lintas di daerah yang telah berkembang dan meningkatkan pelayanan distribusi barang dan jasa guna menunjang pertumbuhan ekonomi. Selain itu, jalan tol memainkan peran penting sebagai bagian dari upaya untuk meningkatkan konektivitas antar kota dan wilayah serta mempercepat mobilitas masyarakat. Banyak manfaat jalan tol yang sudah dirasakan masyarakat Indonesia seperti, jalan tol Jagorawi yang melancarkan lalu lintas sehingga mempersingkat waktu tempuh daerah ke daerah lain, dan masih banyak lagi. Tujuan dari penelitian ini adalah membuat *machine learning* prediksi tarif jalan tol guna memberi acuan kepada masyarakat, mengoptimalkan tarif tol di Indonesia, serta memberikan masukan tarif tol sebagai pertimbangan pemerintah terkait. Pendekatan penelitian ini adalah kuantitatif menggunakan regresi linier dengan algoritma *XGB Regressor*. Hasil pembuatan *machine learning* prediksi tarif tol ini cukup akurat di mana hasil uji akurasi yang menggunakan metrik *root mean squared error* (RMSE) berada di angka 3390.691, dengan hasil *testing* menunjukkan adanya beberapa tarif prediksi yang sesuai dengan tarif asli.

Kata kunci: machine learning, prediksi tarif, regresi linier, tarif tol, *XGB Regressor*.

1 PENDAHULUAN

Jalan tol, yang juga dikenal sebagai jalan bebas hambatan, merupakan infrastruktur vital yang dirancang untuk memfasilitasi lalu lintas kendaraan dengan kecepatan tinggi dan mengurangi kemacetan. Sesuai dengan Peraturan Pemerintah (PP) No. 15 Tahun 2005, jalan tol didefinisikan sebagai jalan umum yang merupakan bagian dari sistem jaringan jalan nasional, di mana pengguna diwajibkan membayar biaya tol untuk menggunakannya.[1] Dalam

beberapa tahun terakhir, perkembangan jalan tol di Indonesia telah mengalami peningkatan yang signifikan, seiring dengan peningkatan kebutuhan akan infrastruktur yang mendukung pertumbuhan ekonomi dan mobilitas masyarakat.

Pertumbuhan ekonomi dan distribusi barang serta jasa sangat bergantung pada efisiensi jaringan transportasi. Jalan tol berperan penting dalam meningkatkan konektivitas antar wilayah, mempercepat mobilitas masyarakat, serta

mendukung distribusi barang dan jasa yang lebih efisien. Contoh konkret dari manfaat jalan tol dapat dilihat pada jalan tol Jagorawi yang telah terbukti mengurangi waktu tempuh antara berbagai daerah, sehingga mendukung aktivitas ekonomi dan mobilitas harian masyarakat. Hingga saat ini, Indonesia memiliki sekitar 2.620 kilometer jalan tol yang beroperasi, yang tersebar di 70 ruas jalan dengan total 481 gerbang tol.[2] Jumlah ini terus bertambah seiring dengan pembangunan jalan tol baru yang diharapkan dapat lebih memperlancar arus lalu lintas dan mendorong pembangunan ekonomi di berbagai wilayah.

Meskipun keberadaan jalan tol memberikan berbagai manfaat, penggunaannya tidak terlepas dari kewajiban membayar biaya tol. Tarif tol merupakan sejumlah uang yang dibayarkan oleh pengguna jalan untuk menggunakan fasilitas ini.[3][4] Dana yang terkumpul dari tarif tol digunakan untuk pengembalian investasi, pemeliharaan, dan pengembangan jalan tol agar infrastruktur ini dapat terus beroperasi dengan baik.[5] Oleh karena itu, penetapan tarif tol menjadi salah satu faktor kunci yang mempengaruhi keberlanjutan penggunaan jalan tol serta aksesibilitas transportasi bagi masyarakat luas.

Penetapan tarif tol di Indonesia didasarkan pada beberapa faktor, di antaranya adalah keuntungan dari operasi kendaraan, kemampuan pengguna jalan, dan nilai investasi yang diperlukan untuk pembangunan dan pemeliharaan jalan tol.[5] Selain itu, tarif tol juga dipengaruhi oleh berbagai fitur seperti ruas jalan, asal perjalanan, tujuan perjalanan, panjang jalan, *latitude*, *longitude*, alamat, gerbang tol, sistem pembayaran, dan golongan kendaraan. Khusus untuk Golongan I, yang mencakup kendaraan seperti Sedan, Jip, *Pick Up*/Truk Kecil, dan Bus, penetapan tarif tol menjadi fokus utama karena tingginya penggunaan golongan ini dalam mobilitas sehari-hari masyarakat.[6]

Dalam konteks modern, penggunaan teknologi *machine learning* menjadi relevan dalam mengatasi kompleksitas penetapan tarif tol. Teknologi ini memungkinkan analisis mendalam terhadap berbagai data dan fitur yang mempengaruhi tarif tol, sehingga dapat memberikan prediksi yang lebih akurat dan optimal. Penelitian ini bertujuan untuk menggabungkan potensi *machine learning* dengan data tarif tol yang ada untuk memprediksi tarif tol Golongan I di jalan tol wilayah Jabodetabek.

Algoritma XGBoost (EXtreme Gradient Boosting) dipilih dalam penelitian ini karena dikenal dengan kinerjanya yang cepat dan kemampuannya dalam menangani *dataset* yang besar. Algoritma yang dikembangkan oleh Tianqi Chen pada tahun 2014 ini merupakan salah satu implementasi lanjutan dari algoritma *gradient boosting* yang efektif untuk masalah regresi, termasuk dalam prediksi tarif tol.[7] XGBoost bekerja dengan membangun beberapa pohon keputusan secara berurutan, di mana setiap pohon bertujuan untuk mengoreksi kesalahan yang dibuat oleh pohon

sebelumnya.[8] Proses ini memungkinkan model untuk memberikan hasil prediksi yang lebih akurat.

Selain XGBoost, algoritma lain yang digunakan untuk tujuan prediksi dalam penelitian ini adalah *Random Forest* dan *Snap Boosting Machine Regression*. *Random Forest* adalah algoritma populer yang digunakan baik untuk tugas klasifikasi maupun regresi, dengan membangun beberapa pohon keputusan secara independen untuk meningkatkan akurasi hasil.[9] *Snap Boosting Machine Regression*, di sisi lain, merupakan algoritma yang membangun *ensemble* dari pohon keputusan tanpa menggunakan kedalaman pohon tetap pada setiap iterasi *boosting*, yang membuatnya fleksibel dan adaptif dalam berbagai kondisi data.[10]

Dengan latar belakang tersebut, penelitian ini akan menjawab beberapa pertanyaan kunci: bagaimana cara membuat model *machine learning* menggunakan regresi linier dengan algoritma XGB *Regressor* untuk memprediksi tarif tol, faktor-faktor apa yang paling berpengaruh dalam prediksi tarif tol di Indonesia, dan sejauh mana akurasi model prediksi yang dihasilkan.

Tujuan Penelitian dari penelitian ini adalah untuk mengembangkan model *machine learning* yang mampu memprediksi tarif tol dengan akurasi tinggi, memahami faktor-faktor yang mempengaruhi tarif tol, serta memberikan acuan bagi penetapan tarif tol yang lebih optimal dan adil.

Manfaat Penelitian ini diharapkan dapat memberikan kontribusi nyata bagi masyarakat dan pemerintah dalam menentukan tarif tol yang sesuai dengan kondisi ekonomi dan kebutuhan mobilitas masyarakat. Penelitian ini juga diharapkan dapat menjadi acuan bagi pengembangan kebijakan tarif tol di masa depan.

Dalam penelitian ini, batasan masalah yang diterapkan meliputi penggunaan dataset tarif tol Golongan I di wilayah Jabodetabek yang bersumber dari data *open source* Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat. Selain itu, uji akurasi akan dilakukan secara otomatis menggunakan algoritma *machine learning* yang telah dijelaskan sebelumnya.

2 METODE PENELITIAN

Pada bagian ini berisi penjelasan tentang jenis penelitian/desain penelitian.

2.1 Metode pengumpulan data, instrumen penelitian, dan metode pengujian

Metode pengumpulan data memerlukan langkah yang strategis dan sistematis dalam penelitian untuk mendapatkan data yang valid dan sesuai dengan kenyataannya. Karena metode pengumpulan data mempunyai tujuan utama yaitu mendapatkan sebuah data. Dalam penelitian kualitatif proses pengumpulan informasi yang dapat diukur dan dinyatakan dalam bentuk angka atau

statistik. Data kuantitatif sangat penting untuk penelitian karena memberikan informasi yang akurat dan terperinci.

Pada penelitian ini menggunakan data sekunder, data sekunder adalah data yang sudah ada, seperti data dari lembaga pemerintah, penelitian sebelumnya, atau basis data publik. Data tarif jalan tol yang digunakan dalam penelitian ini adalah data yang diterbitkan oleh Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat di [website bpjt.pu.go.id](http://website.bpjt.pu.go.id). [11][12] Selain itu ada juga data tarif jalan tol yang sudah berbentuk *file* .CSV berasal dari kaggle.com yang ditulis oleh Muhammad Yusuf Aristyanto.[13]

Namun data tarif jalan tol yang diterbitkan oleh BPJT hanya tarif dari suatu daerah ke daerah yang lain, tidak ada data seperti panjang jalan, gerbang tol, atau *latitude* dan *longitude*. Jadi peneliti harus mengumpulkan data panjang jalan, gerbang tol, *latitude* dan *longitude*; dan data tersebut sangat sulit ditemukan khususnya jalan tol baru seperti jalan tol Depok - Antasari, Cibitung - Cilincing, dll. Sangat jarang sekali ada yang membahas data - data tersebut, hanya ada berita dan itu pun per daerah dan tidak ada *latitude* dan *longitude*, Jadi mengandalkan *google maps* untuk informasi *latitude* dan *longitude*.

Pada pengujian dilakukan uji akurasi pada model menggunakan metrik, dan pada penelitian kali ini metrik yang digunakan adalah *root mean squared error* (RMSE), di mana metrik statistik yang digunakan untuk mengevaluasi kinerja model regresi linier. Secara sederhana, RMSE mengukur rata-rata kesalahan antara nilai aktual dan nilai prediksi yang dihasilkan oleh suatu model. Nilai RMSE yang lebih kecil menunjukkan bahwa model tersebut lebih baik dalam memprediksi hasil. Namun hasil RMSE dari penelitian ini sudah ada otomatis saat data diolah, sehingga penulis tidak mengetahui proses perhitungannya seperti apa. Untuk rumus RMSE adalah sebagai berikut.

$$RMSE = \sqrt{(1/n) * \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Dimana:

- n adalah jumlah data
- y_i adalah nilai sebenarnya
- \hat{y}_i adalah nilai prediksi

Selain uji akurasi dengan metrik *root mean squared error* (RMSE), ada evaluasi model di mana beberapa metrik pengukuran yang terlibat dalam satu model. Semua ditampilkan saat sudah selesai membuat model *machine learning*.

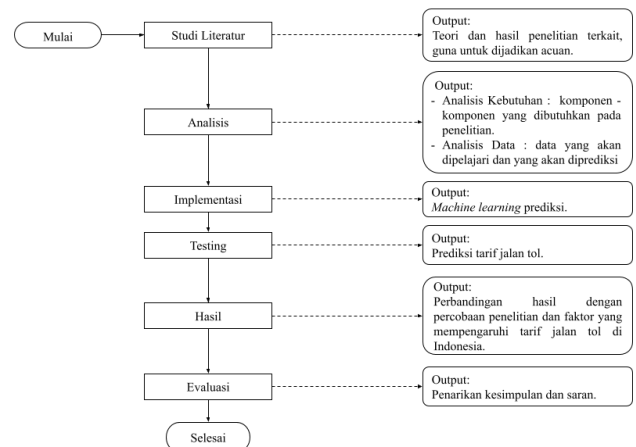
1. *R Squared* merupakan suatu nilai yang memperlihatkan seberapa besar variabel independen mempengaruhi variabel dependen
2. *Explained Variance* (juga disebut *explained variation*) adalah bagian dari total varian model statistik yang

dapat dijelaskan oleh faktor-faktor yang ada dan bukan disebabkan oleh varian kesalahan

3. *Mean Squared Error* adalah istilah dalam statistik yang digunakan untuk mengukur kesalahan rata-rata antara nilai aktual dan nilai prediksi suatu model.
4. *Mean Squared Log Error* adalah metrik yang digunakan dalam masalah regresi untuk mengukur kesalahan prediksi dalam bentuk logaritmik.
5. *Mean Absolute Error* adalah metrik statistik yang digunakan untuk mengukur rata-rata kesalahan absolut antara nilai prediksi model dan nilai aktual.
6. *Median Absolute Error* adalah metrik statistik yang digunakan untuk mengukur akurasi model prediksi dengan menghitung median dari semua perbedaan absolut antara nilai prediksi dan nilai aktual.
7. *Root Mean Square Log Error* adalah metrik statistik yang digunakan untuk mengevaluasi akurasi model prediksi, terutama ketika berurusan dengan data bernilai positif yang memiliki rentang yang luas.

Dan untuk metode penentuan variabel yang berpengaruh menggunakan fitur-fitur yang baru dibuat selama pembuatan model, dan menunjukkan kepentingan relatif setiap prediktor dalam mengestimasi model. Semua itu ditampilkan di menu *feature summary*, di menu *feature summary* bisa melihat semua fitur yang berpengaruh baik itu fitur baru dibuat dan fitur *original* dari data yang diolah. Di menu *feature summary* juga pada bagian *feature importance* atau fitur yang berpengaruh itu angka yang ditunjukkan berupa angka persen jadi memudahkan penulis sebagai pengguna.

2.2 Tahapan penelitian



Gambar 1. Tahapan Penelitian

Penelitian ini dilakukan dengan tahapan - tahapan penelitian pada gambar 1 di atas yang meliputi studi literatur, analisis kebutuhan, analisis data, implementasi, *testing*, hasil, dan evaluasi.

1. Studi Literatur

Pada tahap ini peneliti mengumpulkan teori - teori yang berkaitan dengan penelitian ini. Teori - teori tersebut dijadikan acuan guna memunculkan ide-ide penelitian, dan

mencari metode yang cocok. Dan cara peneliti menemukan studi literatur, yaitu:

1. Pencarian kata kunci
 2. Pencarian subjek
 3. Pencarian di buku dan artikel ilmiah terkini
 4. Pencarian kutipan dalam sumber-sumber ilmiah
 5. Pencarian melalui bibliografi yang diterbitkan
 6. Pencarian melalui sumber orang
 7. Penjelajahan sistematis
2. Analisis Kebutuhan

Pada tahap ini peneliti membuat alur pengerjaan *project* dan menganalisa komponen apa saja yang dibutuhkan untuk membuat *machine learning* prediksi dengan menggunakan servis - servis yang ada di *ibm cloud*. Dan servis - servis yang dipakai adalah

1. *Watson Machine Learning with (Auto AI)*
Watson machine learning with (auto AI) adalah servis dimana *AI* pembuat *machine learning* otomatis berada dan nama fitur tersebut adalah *build machine learning models*.
2. *Watson Studio*
Watson studio adalah servis yang berguna untuk *mendeploy* proyek yang sudah dikerjakan.
3. *Cloud Object Storage*
Cloud object storage adalah servis yang berguna untuk menyimpan proyek - proyek yang sudah dibuat ataupun yang sudah di *deploy*.
4. *Watsonx*
Watsonx adalah sebuah servis yang mewadahi tiga servis lainnya, sehingga bisa menggunakan satu tab saja.

3. Analisis Data

Peneliti mengumpulkan dataset yang akan diolah dan dilatih oleh *machine learning*, yang nantinya akan menjadi sebuah *machine learning* prediksi. *Dataset* yang dikumpulkan bukan hanya data yang akan diolah dan dilatih, namun ada dataset yang menjadi target prediksi. Pada tahap ini peneliti juga menentukan alur pembuatan *machine learning* prediksi dan menentukan referensi penelitian. Serta menentukan metode yang tepat untuk memecahkan masalah utama yang ingin diselesaikan.

Data yang digunakan adalah data tarif jalan tol di Indonesia yang diterbitkan oleh Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat (bpjt.pu.go.id). Untuk *dataset* yang akan diolah dan dilatih adalah data tarif jalan tol yang berada di Jakarta, Bogor, Depok, Tangerang, dan Bekasi (Jabodetabek), karena jalan tol Jabodetabek merupakan jalan tol yang kompleks dan jalan tol pertama di Indonesia berada di Jabodetabek lebih tepatnya jalan tol yang menghubungkan Jakarta dan Bogor. Sedangkan *dataset* target prediksi menggunakan sebagian data jalan tol yang diolah yang mewakili tiap ruas jalan tol.

4. Implementasi

Tahap selanjutnya implementasi dengan menggunakan servis - servis. Pada tahap ini *machine learning* prediksi akan mempelajari *dataset* tarif jalan tol golongan I di Jabodetabek, nantinya pada saat implementasi akan terlihat seberapa akurat prediksi tarif jalan tol golongan I di Indonesia, dan *machine learning* yang sudah dibuat bisa memprediksi tarif tol golongan I di Indonesia dengan catatan format *dataset* yang menjadi target prediksi dengan format *dataset* yang dipelajari sama. Sehingga hasil dari prediksi bisa juga digunakan untuk menganalisa tarif tol golongan I. Untuk mengetahui secara garis besar alur pengerjaan bisa dilihat pada hasil dan pembahasan.

5. Testing

Tahap selanjutnya adalah *testing* dari *machine learning* prediksi yang sudah dibuat. *Testing* ini dilakukan akan memprediksi tarif golongan I dengan fitur yang mempengaruhi yaitu ruas jalan tol, asal perjalanan, tujuan perjalanan, jarak (km), *latitude*, *longitude*, alamat, Gerbang tol, sistem pembayaran, dan tarif. Tujuan pada tahap pengujian ini adalah untuk mengevaluasi apakah *machine learning* prediksi yang sudah dibuat bisa memprediksi sesuai apa yang diharapkan. Pada tahap ini terbagi menjadi dua tahap, pada tahap pertama untuk mendapatkan nilai acuan selisih hasil prediksi, lalu tahap kedua memprediksi target tujuan penelitian ini.

6. Hasil

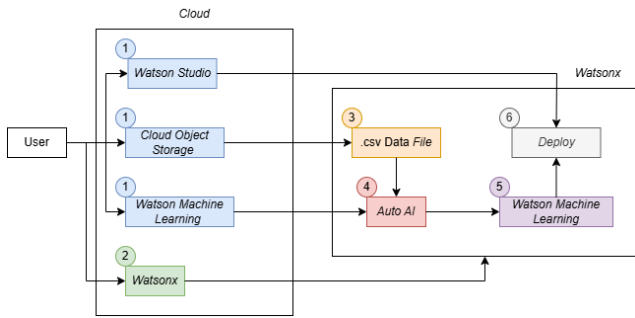
Di tahap ini menjelaskan cara membuat *machine learning* prediksi dengan aplikasi *ibm cloud*, yang nantinya akan terlihat seberapa besar pengaruh variabel dengan hasil. Di tahap ini juga membandingkan *testing* tahap dua dengan data asli. Selain itu hasil dari perbandingan *testing* yang nantinya akan menghasilkan apakah data yang dihasilkan sesuai yang diharapkan, dan suatu kesimpulan dari rumusan masalah yang ada.

7. Evaluasi

Tahap yang terakhir adalah tahap evaluasi dari keseluruhan tahapan penelitian. Pada tahap ini peneliti mengevaluasi apa saja kekurangan dalam penelitian ini, dan kekurangan tersebut akan dijadikan saran pada penelitian berikutnya.

3 HASIL DAN PEMBAHASAN

3.1 Alur Pengerjaan



Gambar 2. Alur Pengerjaan

Berikut merupakan penjelasan alur pengerjaan berdasarkan gambar 2:

- 1 User menambahkan servis *cloud object storage*, *watson wtudio*, *watson machine learning*, dan *watsonx* sebagai persyaratan untuk membuat *machine learning*.
- 2 User mengakses *watsonx* yang sudah terintegrasi otomatis oleh *cloud object storage*, *watson studio*, dan *watson machine learning*.
- 3 User mengunggah Jabodetabek toll road rates dalam format file .csv ke dalam *watsonx*.
- 4 User membuat percobaan *watson machine learning with (auto AI)* untuk memprediksi tarif tol di Jabodetabek di *watsonx*
- 5 *Auto ai* menggunakan *watson machine learning* untuk membuat beberapa model, dan user menerapkan model dengan performa terbaik.
- 6 User mendeploy model dengan performa terbaik dengan *watson studio* di *watsonx*.

3.2 Hasil

Pada sub bab ini penulis akan memaparkan variabel apa yang mempengaruhi hasil prediksi, hasil *testing* proyek, dan membandingkan data hasil prediksi dengan data asli. Dan penulis melakukan percobaan sebanyak 3 kali, di mana percobaan pertama dilakukan pada saat kegiatan Studi Independen, percobaan kedua dilakukan pada bulan Mei 2024 minggu kedua hingga minggu ketiga, Dan percobaan ketiga dilakukan pada bulan Juni 2024 minggu pertama.

Dari ketiga percobaan yang dilakukan percobaan ketiga adalah yang paling akurat dengan nilai RMSE di bawah 5000. Percobaan ketiga menghasilkan tiga algoritma dan hasil dari ketiga algoritma tersebut yang dapat dilihat pada tabel 1 berikut.

Tabel 1. Hasil Penelitian

Algoritma	<i>Extreme Gradient Boosting Regressor</i>	<i>Snap Boosting Machine Regressor</i>	<i>Random Forest Regressor</i>
Data	Jabodetabek	Jabodetabek	Jabodetabek
Akurasi	Root Mean Square Error = 3390.691	Root Mean Square Error = 4027.684	Root Mean Square Error = 4554.243
Fitur Berpengaruh	Asal Pejalan 20,14%, Panjang Jalan 3,43%, dan Tujuan Perjalanan 0,89%	Asal Perjalanan 14.23%, Panjang Jalan 8,63% dan Ruas Jalan Tol 4,36%	Panjang Jalan 31,76%, Asal Perjalanan 10,11%, Latitude 7,51%, dan Longitude 7,26%
Hasil Prediksi	Sangat baik, ada 7 yang sesuai tarif asli dengan tarif prediksi dengan selisih 500 - 15500	Sangat baik, ada 7 yang sesuai tarif asli dengan tarif prediksi dengan selisih 500 - 15500	Baik, ada 3 yang sesuai dengan tarif asli dengan prediksi, dengan selisih 500 - 9000.

Pada algoritma *extreme gradient gosting regressor* akurasi yang ditunjukkan oleh metrik *root mean square error* sebesar 3390.69. Fitur berpengaruh ada asal perjalanan sebesar 20,14%, panjang jalan sebesar 3,43%, dan tujuan perjalanan sebesar 0,89%. Dan hasil prediksi sangat baik, ada 7 yang sesuai tarif asli dengan tarif prediksi, dengan selisih sebesar 500 - 15500.

Pada algoritma *snap boosting machine regressor* akurasi yang ditunjukkan oleh metrik *root mean square error* sebesar 4027.684. Fitur berpengaruh ada asal perjalanan sebesar 14.23%, lalu panjang jalan sebesar 8,63%, dan terakhir ruas jalan tol mempengaruhi sebesar 4,36%. Dan pada dasarnya hasil prediksi tidak jauh berbeda dengan algoritma *XGB Regressor* di mana ada 7 yang sesuai tarif asli dengan tarif prediksi, dengan selisih sebesar 500 - 15500.

Pada algoritma *random forest regressor* akurasi yang ditunjukkan oleh metrik *root mean square error* sebesar 4027.684. Fitur berpengaruh ada panjang jalan sebesar 31,76%, kemudian asal perjalanan sebesar 10,11%, lalu *latitude* sebesar 7,51%, dan terakhir *longitude* mempengaruhi sebesar 7,26%. Dan hasil prediksi baik, ada 3 yang sesuai tarif asli dengan tarif prediksi, dengan selisih 500 - 9000.

Dari poin - poin yang dijelaskan algoritma yang paling akurat dari ketiga algoritma tersebut adalah *extreme gradient gosting regressor* dengan *root mean square error* di angka 3390.691, namun hasil *testing* selisih tarif asli dengan tarif prediksi masih lebih baik pada algoritma *random forest regressor*. Tetapi jika hasil *testing* selisih tarif asli dengan tarif prediksi yang terbaik adalah algoritma *snap boosting machine regressor*, namun jika melihat pengaruh fitur lainnya yang paling sesuai adalah algoritma *extreme gradient gosting regressor*. Dimana pada algoritma *extreme gradient gosting regressor* hasil tarif asli dengan tarif prediksi yang jauh, terpengaruh oleh fitur lainnya.

4 KESIMPULAN

1. *Machine learning* prediksi yang menggunakan regresi linier dengan algoritma *XGB Regressor* untuk memprediksi tarif tol dapat dibuat dengan mempersiapkan data yang matang, memilih model yang tepat sesuai dengan referensi yang diprediksi, melatih model, dan melakukan evaluasi model menggunakan metrik.
2. Faktor atau fitur yang paling berpengaruh dalam memprediksi tarif tol di Indonesia yang paling optimal, di mana asal perjalanan di angka 20,14%, panjang jalan di angka 3,43%, dan tujuan perjalanan di angka 0,89%.
3. Pada penelitian selanjutnya disarankan untuk melakukan pembahasan cara mengimplementasikan model *machine learning* tersebut menjadi berbasis *website*.

Ucapan Terima Kasih

Penulis S.A.K mengucapkan terima kasih kepada dosen pembimbing dan penguji yang telah memberikan masukan serta saran-saran. Penulis juga mengucapkan terima kasih kepada Infintie Learning Indonesia yang bekerja sama dengan IBM yang telah menyediakan layanan servis yang dipakai dalam penelitian ini.

DAFTAR PUSTAKA

- [1] Indonesia, "Peraturan Pemerintah (PP) Nomor 15 Tahun 2005 Tentang Jalan Tol," 2005.
- [2] Sistem Informasi Geografis Badan Pengatur Jalan Tol, "Peta Jalan Tol Indonesia," bpjt.pu.go.id. Accessed: Mar. 14, 2024. [Online]. Available: <https://sigi.pu.go.id/portalpupr/apps/dashboards/ad691982b770462d8e236f8ca7e450f4>
- [3] Indonesia, "Peraturan Pemerintah (PP) Nomor 30 Tahun 2017 Tentang Perubahan Ketiga atas Peraturan Pemerintah Nomor 15 Tahun 2005 tentang Jalan Tol," 2017.
- [4] Rosalita L, Purba A, and Sulistiyorini R, "Analisis Tarif Tol Terbanggi Besar – Pematang Panggang Berdasarkan Kemauan Membayar dan Kemampuan Membayar Masyarakat," *Jurnal Teknik Sipil*, vol. 7, no. 2, Jun. 2019.
- [5] Indonesia, "Undang - Undang Republik Indonesia Nomor 38 Tahun 2004 Tentang Jalan," 2004.
- [6] Indonesia, "Peraturan Menteri Pekerjaan Umum dan Perumahan Rakyat (PUPR) No. 16/PRT/M/2017," 2017.
- [7] B. Jange, "Prediksi Harga Saham Bank BCA Menggunakan XGBoost," *ARBITRASE: Journal of Economics and Accounting*, vol. 3, no. 2, pp. 231–237, Nov. 2022, doi: 10.47065/arbitrase.v3i2.495.
- [8] J. Avanija, G. Sunitha, K. Reddy Madhavi, P. Kora, and R. Vittal Sai Hitesh, "Prediction of House Price Using XGBoost Regression Algorithm," 2021.
- [9] S. Zahedian, A. Nohekhan, and K. F. Sadabadi, "Dynamic toll prediction using historical data on toll roads: Case Study of the I-66 Inner Beltway," *Transportation Engineering*, vol. 5, Sep. 2021, doi: 10.1016/j.treng.2021.100084.
- [10] T. Parnell *et al.*, "SnapBoost: A Heterogeneous Boosting Machine".
- [11] Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat, "Tarif Tol," bpjt.pu.go.id. Accessed: Feb. 03, 2024. [Online]. Available: <https://bpjt.pu.go.id/tabel-tarif-tol>
- [12] Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat, "Cek Tarif Tol," bpjt.pu.go.id. Accessed: Feb. 03, 2024. [Online]. Available: <https://bpjt.pu.go.id/cek-tarif-tol>
- [13] M. Yusuf Aristyanto, "Indonesian Toll Road Rates in June 2020," kaggle.com.