



SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI

**PEMANFAATAN *MACHINE LEARNING* UNTUK PREDIKSI
TARIF TOL MENGGUNAKAN ALGORITMA *XGB*
*REGRESSOR***

TUGAS AKHIR

SAID AL KHAIRI

0110220227

**PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI
DEPOK
AGUSTUS 2024**



**STT TERPADU
NURUL FIKRI**

SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI

**PEMANFAATAN *MACHINE LEARNING* UNTUK PREDIKSI
TARIF TOL MENGGUNAKAN ALGORITMA *XGB*
*REGRESSOR***

TUGAS AKHIR

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
Komputer**

STT - NF

**SAID AL KHAIRI
0110220227**

**PROGRAM STUDI TEKNIK INFORMATIKA
SEKOLAH TINGGI TEKNOLOGI TERPADU NURUL FIKRI
DEPOK
AGUSTUS 2024**

HALAMAN PERNYATAAN ORISINALITAS

Skripsi/Tugas Akhir ini adalah hasil karya penulis, dan semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar.

Nama : Said Al Khairi

NIM : 0110220277

Bogor, 11 Agustus 2024

STT - NF Tanda Tangan



Said Al Khairi

HALAMAN PENGESAHAN

Tugas Akhir ini diajukan oleh :

Nama : Said Al Khairi

NIM : 0110220277

Program Studi : Teknik Informatika

Judul Skripsi : Pemanfaatan *Machine Learning* Untuk Prediksi Tarif Tol
Menggunakan Algoritma *XGB Regressor*

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Sekolah Tinggi Teknologi Terpadu Nurul Fikri.

DEWAN PENGUJI

Pembimbing

Penguji



Ahmad Rio Adriansyah, S.Si., M.Si.



Dr. Lukman Rosyidi, M.T., M.M.

STT - NF

Ditetapkan di : Depok

Tanggal : 29 Juli 2024

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT, karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan skripsi/Tugas Akhir ini. Penulisan skripsi/Tugas Akhir ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana komputer Program Studi Teknik Informatika pada Sekolah Tinggi Teknologi Terpadu Nurul Fikri Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi penulis untuk menyelesaikan skripsi/tugas akhir ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Allah SWT.
2. Orang tua dan semua anggota keluarga yang telah memberikan dorongan baik secara moril maupun materil dalam penyelesaian tugas ini.
3. Bapak Dr. Lukman Rosyidi, M.T., M.M. selaku Ketua Sekolah Tinggi Teknologi Terpadu Nurul Fikri dan selaku Dosen Penguji Tugas Akhir penulis dalam menyelesaikan penulisan ilmiah ini..
4. Ibu Nurul Janah, S.IIP., M.Hum selaku Dosen Pembimbing Akademik yang telah membimbing penulis selama berkuliah di Sekolah Tinggi Teknologi Terpadu Nurul Fikri.
5. Bapak Ahmad Rio Adriansyah, S.Si., M.Si. selaku Dosen Pembimbing Tugas Akhir penulis dalam menyelesaikan penulisan ilmiah ini.
6. Sahabat dan teman - teman yang telah banyak membantu saya dalam menyelesaikan skripsi/tugas akhir ini.

Akhir kata, penulis berharap Allah SWT berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga skripsi ini membawa manfaat bagi pengembangan ilmu.

Bogor, 11 Agustus 2024



Penulis

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Sekolah Tinggi Teknologi Terpadu Nurul Fikri, saya yang bertanda tangan di bawah ini:

Nama : Said Al Khairi

NIM : 0110220277

Program Studi : Teknik Informatika

Jenis karya : Tugas Akhir

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada STT- NF **Hak Bebas Royalti Noneksklusif** (*Non-exclusive Royalty - Free Right*) atas karya ilmiah saya yang berjudul :

Pemanfaatan *Machine Learning* Untuk Prediksi Tarif Tol Menggunakan Algoritma *XGB Regressor*

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini STT-NF berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di: Bogor

Pada tanggal : 11 Agustus 2024

Yang menyatakan



Said Al Khairi

ABSTRAK

Nama : Said Al Khairi

Program Studi : Teknik Informatika

Judul : Pemanfaatan *Machine Learning* Untuk Prediksi Tarif Tol
Menggunakan Algoritma *XGB Regressor*

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 dimana 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 : tarif tol, *machine learning*, *xgb regressor*.

STT - NF

ABSTRACT

Name : Said Al Khairi

Study Program : Informatics Engineer

Title : *Utilisation Of Machine Learning For Toll Rates Prediction
Using XGB Regressor Algorithm*

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.

Key words : toll rates, machine learning, xgb regressor.

STT - NF

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	iii
HALAMAN PENGESAHAN	iv
KATA PENGANTAR	v
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	2
1.5 Batasan Masalah	3
1.6 Sistematika Penulisan	3
BAB II KAJIAN LITERATUR	5
2.1. Jalan Tol	5
2.2. Tarif Tol	6
2.3. Regresi Linier	7
2.4. XGB Regressor	8
2.5. Snap Boosting Machine Regressor	9
2.6. Random Forest	10
2.7. IBM Cloud	10
2.8. Dataset	11
2.9. Penelitian Terkait	12
BAB III METODOLOGI PENELITIAN	16
3.1. Tahapan Penelitian	16
3.1.1. Studi Literatur	17
3.1.2. Analisis Kebutuhan	17
3.1.3. Analisis Data	18
3.1.4. Implementasi	18
3.1.5. Testing	19
3.1.6. Hasil	19

3.1.7. Evaluasi.....	19
3.2. Jenis Penelitian	19
3.3. Metode Analisis Data.....	20
3.4. Metode Pengumpulan Data.....	21
3.5. Metode Pengujian	22
BAB IV IMPLEMENTASI DAN HASIL.....	24
4.1. Alur Pengerjaan.....	24
4.2. Implementasi	25
4.3. Hasil	43
4.3.1. Percobaan 1	43
4.3.2. Percobaan 2	44
4.3.3. Percobaan 3	45
4.3.3.1. <i>XGB Regressor</i>	46
4.3.3.2. <i>Snap Boosting Machine Regressor</i>	47
4.3.3.3. <i>Random Forest Regressor</i>	48
BAB V KESIMPULAN DAN SARAN	51
5.1 Kesimpulan.....	51
5.2 Saran.....	51
DAFTAR PUSTAKA.....	52
LAMPIRAN.....	54
Lampiran 1	54
Lampiran 2	57
Lampiran 3	60

STT - NF

DAFTAR GAMBAR

Gambar 1 Peta Proyek Jalan Tol Di Jabodetabek	5
Gambar 2 Tahapan Penelitian	16
Gambar 3 Alur Pengerjaan.....	24
Gambar 4 Dataset Jabodetabek Toll Road Rates (Gol I).....	25
Gambar 5 Resource List.....	26
Gambar 6 Watsonx	26
Gambar 7 Home Service Watsonx.....	27
Gambar 8 Projects.....	28
Gambar 9 Start Projects	28
Gambar 10 Define Projects	29
Gambar 11 Upload Dataset.....	29
Gambar 12 Konfigurasi Model Machine Learning.....	30
Gambar 13 Konfigurasi Model Machine Learning.....	30
Gambar 14 Konfigurasi Model Machine Learning.....	31
Gambar 15 Prediction Type	31
Gambar 16 Algorithms To Include	32
Gambar 17 Algorithms To Use.....	32
Gambar 18 Proses Run Experiment	33
Gambar 19 Hasil Experiment.....	34
Gambar 20 Model Evaluation.....	35
Gambar 21 Feature Summary	35
Gambar 22 Save As Untuk Deployment.....	36
Gambar 23 Promote To Deployment	36
Gambar 24 Create New Deployment	37
Gambar 25 Create New Deployment.....	37
Gambar 26 Create New Deployment.....	38
Gambar 27 Promote Deployment	38
Gambar 28 Promote Deployment	39
Gambar 29 Deploy Project.....	39
Gambar 30 Deploy Project.....	40
Gambar 31 Manage Deployment	40
Gambar 32 Deploy Project.....	41
Gambar 33 Prediksi Tarif Tol	42
Gambar 34 Testing Project	42
Gambar 35 Data Percobaan 1	54
Gambar 36 Hasil Percobaan 1.....	54
Gambar 37 Feature Summary Percobaan 1.....	55
Gambar 38 Model Evaluation	55
Gambar 39 Data Percobaan 2.....	57
Gambar 40 Hasil Percobaan 2.....	57
Gambar 41 Feature Summary Percobaan 2.....	58
Gambar 42 Model Evaluation	58
Gambar 43 Feature Summary Percobaaan 3 (Snap Boosting Machine Regressor)	62
Gambar 44 Feature Summary Percobaaan 3 (Random Rofest Regressor)	64

DAFTAR TABEL

Tabel 1 Penelitian Terkait.....	12
Tabel 2 Perbandingan Ketiga Algoritma	49
Tabel 3 Data Hasil Prediksi Percobaan 1	55
Tabel 4 Data Hasil Prediksi Percobaan 2.....	58
Tabel 5 Data Hasil Prediksi Percobaan 3 (XGB Regressor).....	60
Tabel 6 Data Hasil Prediksi Percobaan 3 (Snap Boosting Machine Regressor)....	62
Tabel 7 Data Hasil Prediksi Percobaan 3 (Random Forest Regressor)	64



STT - NF

DAFTAR LAMPIRAN

Lampiran 1	54
Lampiran 2	57
Lampiran 3	60



STT - NF

BAB I

PENDAHULUAN

1.1 Latar Belakang

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.

Jalan tol mewajibkan penggunanya membayar tol sesuai tarif, uang itu digunakan untuk pengembalian investasi, pemeliharaan dan pengembangan jalan tol.[1] Oleh karena itu tarif tol yang dikenakan kepada pengguna adalah faktor kunci dalam keberlanjutan penggunaan jalan tol dan juga mempengaruhi aksesibilitas transportasi bagi masyarakat luas. Penetapan tarif tol didasarkan pada tiga faktor: besarnya keuntungan dari operasi kendaraan, kemampuan pengguna jalan, dan nilai investasi.[1]

Penetapan tarif tol tidak semata-mata bergantung pada pertimbangan biaya operasional jalan tol semata, tetapi juga dipengaruhi oleh faktor - faktor seperti asal perjalanan, tujuan perjalanan, panjang jalan, dan golongan kendaraan. Penetapan tarif tol untuk golongan 1 menjadi fokus utama karena penggunaannya yang luas dalam mobilitas sehari-hari masyarakat. Golongan I dalam golongan kendaraan, yang meliputi kendaraan seperti Sedan, Jip, Pick Up/Truk Kecil, dan Bus.[2] Dalam konteks ini, penggunaan *machine learning* menjadi relevan dalam mengatasi tantangan tersebut. Teknologi *machine learning* memungkinkan analisis yang lebih mendalam terhadap data faktor - faktor yang mempengaruhi tarif tol, yang pada gilirannya dapat digunakan untuk membuat prediksi tarif tol yang lebih akurat. Dan

pemanfaatan *machine learning* juga mengoptimalkan penetapan tarif tol untuk memprediksi tarif tol golongan I di jalan tol Jabodetabek. Oleh karena itu, penelitian "Pemanfaatan *Machine Learning* Untuk Prediksi Tarif Jalan Tol Menggunakan Algoritma *XGB Regressor*" yang menggunakan servis dari *ibm cloud* diharapkan mampu untuk menggabungkan potensi *machine learning* dengan data tarif tol yang berguna memberikan jawaban sementara dan memberi acuan tarif tol golongan I jalan tol di Indonesia.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, masalah yang akan dicari solusinya dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara membuat *machine learning* yang menggunakan regresi linier dengan algoritma *xgb regressor* untuk memprediksi tarif tol?
2. Faktor atau fitur apa yang paling berpengaruh dalam memprediksi tarif tol di Indonesia?

1.3 Tujuan Penelitian

1. Membuat *machine learning* untuk memprediksi tarif tol
2. Memprediksi tarif tol menggunakan regresi linier dengan algoritma *xgb regressor*
3. Mengetahui apa - apa yang mempengaruhi hasil prediksi selain faktor atau fitur yang mempengaruhi tarif tol di Indonesia.
4. Mengetahui pengaruh terbesar faktor atau fitur yang mempengaruhi tarif tol di Indonesia.
5. Mengetahui seberapa jauh perbedaan tarif asil dengan tarif yang di prediksi.

1.4 Manfaat Penelitian

1. Mengoptimalkan tarif tol di Indonesia
2. Memberikan masukan tarif tol sebagai pertimbangan pemerintah terkait.

1.5 Batasan Masalah

Dari identifikasi masalah tersebut, maka dalam penelitian ini terdapat beberapa batasan masalah pada sebagai berikut:

1. Komponen dalam memprediksi tarif tol di Indonesia yaitu ruas jalan tol, asal perjalanan, tujuan perjalanan, jarak (km), *latitude*, *longitude*, alamat, gerbang tol, sistem pembayaran, dan golongan kendaraan.
2. Dataset yang digunakan merupakan dataset open source dari Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat (bpjt.pu.go.id).
3. Uji akurasi dilakukan secara otomatis saat dataset diolah.
4. Dataset yang akan diolah dibatasi hanya data tarif jalan tol golongan I di Jabodetabek.

1.6 Sistematika Penulisan

Untuk memberi gambaran yang jelas tentang penelitian ini, maka disusunlah sistematika penulisan yang berisi materi yang akan dibahas pada setiap bab.

Sistematika dalam penulisan tugas akhir ini adalah sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini menjelaskan tentang latar belakang dari topik penulisan tugas akhir, pokok permasalahan berupa rumusan masalah dan dibatasi dengan batasan masalah, serta tujuan dan manfaat yang diharapkan dari penulisan tugas akhir.

BAB II TINJAUAN PUSTAKA

Pada bab ini menjelaskan seluruh landasan teori yang berhubungan dengan penelitian, yaitu tentang variabel yang digunakan, alat yang digunakan, metode dan algoritma yang digunakan, serta perbandingan penelitian terkait.

BAB III METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang tahapan penelitian yang mana menjelaskan tentang alur penelitian, jenis penelitian, metode analisis data, metode pengumpulan data digunakan beserta sumber data, dan metode pengujian yang digunakan.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan tentang alur pengerjaan, Implementasi, serta hasil percobaan yang dilakukan.

BAB V KESIMPULAN DAN SARAN

Bab ini merupakan penutup yang berisi kesimpulan dan saran dari penelitian yang sudah dilakukan.



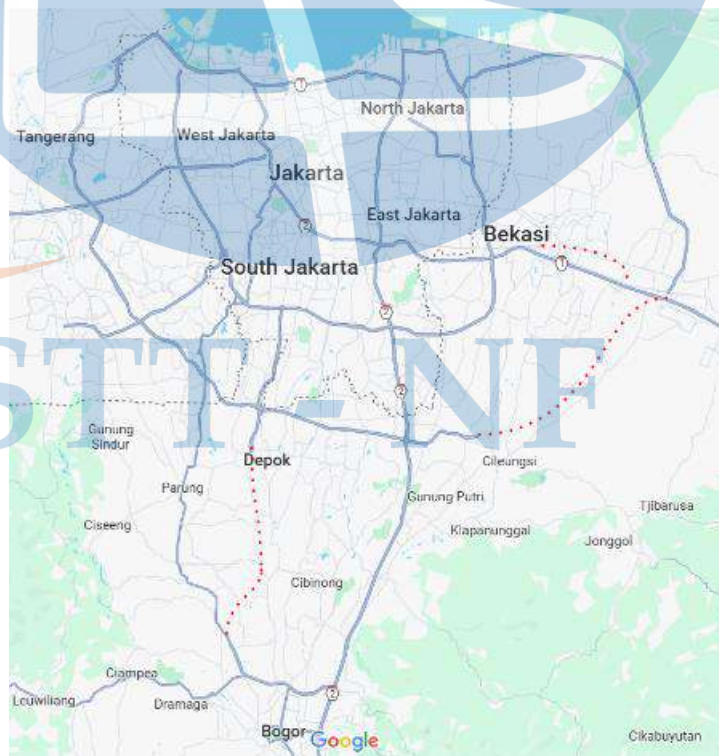
STT - NF

BAB II KAJIAN LITERATUR

2.1. Jalan Tol

Jalan tol, atau yang juga dikenal sebagai jalan bebas hambatan, adalah jenis jalan yang dirancang khusus untuk memfasilitasi lalu lintas kendaraan dengan kecepatan tinggi. Menurut PP No 15 Tahun 2005 jalan tol adalah jalan umum yang merupakan bagian sistem jaringan jalan dan sebagai jalan nasional yang penggunaannya diwajibkan membayar tol.[3] Melansir dari peta jalan tol Indonesia saat ini kurang lebih 2.620 km jalan tol yang beroperasi, serta 70 ruas jalan, dan 481 gerbang tol.[4]

Dan saat ini ada ruas jalan tol Jabodetabek yang sedang dibangun seperti : Ruas jalan Bekasi - Cawang - Kampung Melayu seksi 2B : Hasibuan - Tambun sepanjang 6,90 km, Ruas jalan Depok - Antasari seksi 3 : Sawangan - Bojonggede sepanjang 9,50 km dan seksi 4 : Bojonggede - Salabenda sepanjang 6,40, Ruas jalan Cimanggis - Cibitung seksi 2B : : Cikeas - Cibitung sepanjang 19,80 km, dan masih banyak lagi.[5]



Gambar 1 Peta Proyek Jalan Tol Di Jabodetabek

2.2. Tarif Tol

Tarif tol adalah sejumlah uang tertentu yang dibayarkan untuk penggunaan jalan tol.[6][7] Tarif tol dihitung berdasarkan kemampuan bayar pengguna jalan, besar keuntungan biaya operasi kendaraan, dan kelayakan investasi.[1] Berdasarkan faktor tersebut setidaknya ada lima variabel yang menjadi dasar penentuan tarif jalan tol.

1. Ruas Jalan Tol

Ruas jalan tol sering kali berada di lokasi yang strategis, dimana daerah yang dihubungkan banyak masyarakat tuju atau ekonomi daerah sedang naik. Dan tujuan utama jalan tol adalah mempersingkat waktu, ada beberapa ruas jalan tol yang membelah bukit seperti jalan tol Cipali, jalan tol Semarang - Solo, dsb. Lalu ada juga jalan tol diatas laut yang melewati sebuah teluk, seperti jalan tol Bali Mandara.

2. Asal Perjalanan & Tujuan Perjalanan

Asal Perjalanan & Tujuan Perjalanan menjadi salah satu faktor yang penentu tarif jalan tol, dimana setiap daerah memiliki nilai investasi yang berbeda - beda sehingga tarif jalan tol setiap daerah berbeda juga. Dan tarif ini ditentukan oleh Menteri Pekerjaan Umum dan Perumahan Rakyat bersama Badan Pengatur Jalan Tol.[1][3]

3. Panjang Jalan/Jarak Tempuh

Panjang jalan menjadi kunci utama dalam penentuan tarif jalan tol, hal ini terjadi karena semakin dekat jarak tempuh maka tarif yang dikenakan juga semakin murah. Walaupun ada beberapa ruas jalan tol yang tarifnya flat, jadi jauh dekat pembayarannya sama. Hal ini terjadi dikarenakan sistem transaksi terbuka seperti tol Dalam Kota, tol Lingkar Luar, dan tol Jagorawi.

4. Gerbang Tol

Gerbang tol adalah bagian dari jalan tol yang berfungsi sebagai gerbang masuk dan keluar dari jalan tol dan tempat untuk mengumpulkan tarif tol dari pengguna jalan tol. Gerbang tol sendiri memiliki titik yang jelas di peta, sehingga memiliki alamat dan titik *latitude*, *longitude*. Karena alamat dan titik *latitude*, *longitude* inilah gerbang tol memiliki nilai yang mempengaruhi tarif tol.

5. Golongan Kendaraan

Jalan tol adalah jalan umum maka dari itu pengguna jalan tol tidak hanya mobil sedan, jip, atau bus, ada juga truk - truk besar seperti, truk dengan dua gandar, atau 3 gandar, dll.[8] Oleh karena itu, golongan kendaraan juga menjadi salah satu faktor yang penentu tarif jalan tol.

2.3. Regresi Linier

Regresi linier adalah suatu metode yang biasanya untuk menemukan persamaan dari sebuah data yang dimana data tersebut saling berhubungan antara variabel satu dengan variabel lain yang ingin diramalkan dalam satu database yang cukup besar.[9] Metode regresi linier juga menjelaskan hubungan antara dua variabel yakni, variabel terikat (Y) dan variabel bebas (X).[10][11] Sebuah variabel hasil observasi yang diperoleh sangat mungkin dipengaruhi oleh variabel lainnya, misalkan tinggi badan dan berat badan seseorang.[12] banyak digunakan dalam berbagai bidang, termasuk bidang ekonomi dan industri, antara lain. Dalam bidang ekonomi, regresi dapat digunakan untuk memprediksi jumlah pengeluaran bulanan tiap keluarga dibandingkan dengan penghasilan mereka. Di bidang industri, regresi dapat digunakan untuk menentukan dan memprediksi harga bahan baku industri terhadap ketersediaan bahan baku industri di pasar.[13] Dan di bidang farmasi dapat memprediksi persediaan obat, dan masih banyak lagi kegunaan metode regresi ini. Namun metode regresi ini dibagi menjadi 3 macam yaitu:

- a. **Regresi Linear Sederhana** menggunakan data kuantitatif untuk melihat hubungan antara dua variabel.[13]
- b. **Regresi Linear Berganda** menggunakan data kuantitatif untuk menghubungkan satu variabel Y ke dua atau lebih variabel X.[13]
- c. **Regresi Non Linear** menggunakan data kuantitatif untuk menghubungkan variabel tidak linear yang disebut X dan Y.[13]

Dan penelitian ini menggunakan regresi linier berganda atau *multiple linear regression* karena data yang digunakan pada penelitian ini ada lebih dari satu variabel yang menghubungkan variabel lain. Untuk rumus regresi linier berganda sebagai berikut.

$$Y = a + b_1X_1 + b_2X_2 + \dots + e$$

$$b_1 = \frac{(\sum X_2^2)(\sum X_2 \sum Y) - (\sum X_1 \sum X_2)(\sum X_2 \sum Y)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 \sum X_2)^2}$$

$$b_2 = \frac{(\sum X_1^2)(\sum X_1 \sum Y) - (\sum X_1 \sum X_2)(\sum X_1 \sum Y)}{(\sum X_1^2)(\sum X_2^2) - (\sum X_1 \sum X_2)^2}$$

$$a = \frac{\sum Y}{n} - \frac{b_1 \sum X_1}{n} - \frac{b_2 \sum X_2}{n}$$

Keterangan:

a : konstanta (titik potong Y)

b : koefisien dari variabel X (koefisien determinasi)

Y : variabel dependen

X : variabel independen

e : error

Namun penelitian ini tidak berfokus pada perhitungan prediksi tetapi berfokus pada analisis hasil apakah tarif jalan tol bisa diprediksi menggunakan regresi linier. Lalu variabel mana saja yang berpengaruh terhadap hasil prediksi.

2.4. *XGB Regressor*

XGBoost, juga dikenal sebagai *extreme gradient boosting*, adalah algoritma pembelajaran kelompok dengan metode peningkatan yang dikembangkan oleh *Tianqi Chen* pada tahun 2014. Algoritma ini didasarkan pada prinsip peningkatan gradient.[14] Dimana algoritma tersebut bukan hanya digunakan untuk masalah regresi, tetapi bisa digunakan juga pada deret waktu, seperti memprediksi hasil yang berkelanjutan seperti harga rumah, harga saham, dll.[15][14] *XGBoost* adalah implementasi lanjutan dari algoritma gradient boosting yang dikenal dengan kecepatan dan kinerjanya dalam menangani set data yang besar.[15] Algoritma ini bekerja dengan membangun beberapa pohon keputusan secara berurutan, dengan

setiap pohon mengoreksi kesalahan dari pohon sebelumnya, sehingga menghasilkan model prediksi yang lebih akurat.[15]

XGBoost dikenal sebagai algoritma yang efisien dan fleksibel, sehingga banyak digunakan dalam berbagai permasalahan regresi di dunia nyata. Selain itu berikut adalah kelebihan *xgboost*.

1. Akurasi yang tinggi: *xgboost* seringkali bisa menghasilkan model prediksi yang sangat akurat, terutama untuk data yang kompleks.
2. Dapat menangani berbagai jenis data: *xgboost* bisa digunakan untuk data numerik maupun kategorikal.
3. Relatif cepat: Beberapa implementasi *xgboost* dirancang untuk berjalan dengan efisien, terutama pada perangkat keras modern.

2.5. *Snap Boosting Machine Regressor*

Algoritma *Snap Boosting Machine Regression* adalah algoritma pembelajaran mesin yang digunakan untuk membangun ensemble dari pohon keputusan.[16] Algoritma ini dapat digunakan baik untuk tugas klasifikasi maupun regresi. Yang membedakan boosting machine dari *framework boosting* lainnya adalah bahwa *Snap boosting machine* tidak menggunakan *fixed maximal tree depth* pada setiap iterasi boosting.[16] Cara kerja algoritma ini adalah dengan mengambil sampel subkelas secara acak pada setiap iterasi boosting dan memilih hipotesis dasar yang meminimalkan fungsi kerugian untuk subkelas yang diambil sampelnya.[16] Berikut adalah rumus dari algoritma *snap boosting machine regressor*.

$$\min_{f \in \mathcal{F}} \sum_{i=1}^n l(y_i, f(x_i)), \quad \text{STT - NF}$$

Seperti yang dijelaskan pada sub bab 2.3. dimana penelitian ini tidak berfokus pada perhitungan prediksi tetapi berfokus pada analisis hasil, dikarenakan penelitian ini menggunakan *IBM Cloud* sebagai aplikasi.

2.6. *Random Forest*

Random Forest adalah algoritma pembelajaran mesin yang populer digunakan untuk tugas klasifikasi dan regresi. Algoritma ini terdiri dari banyak pohon keputusan (decision trees) untuk mencapai hasil yang lebih akurat.[17] Algoritma ini bekerja dengan membangun beberapa pohon keputusan secara independen.[17] Kumpulan pohon ini membentuk sebuah model pembelajaran yang meningkatkan akurasi hasil. Setiap model dilatih secara terpisah hingga mengeluarkan hasil yang bisa merepresentasikan atau memprediksi seluruh model. Variabel untuk memisahkan data di setiap node adalah ditentukan dengan menggunakan kriteria *Gini Impurity* seperti yang ditunjukkan pada persamaan dibawah ini.

$$Gini\ Impurity = 1 - (P_1)^2 - (P_2)^2$$

dimana P_i , adalah probabilitas observasi yang termasuk dalam kelas i . Variabel dengan "*Impurity*" terendah adalah variabel yang dipilih untuk memisahkan data.

Namun seperti yang dijelaskan pada sub bab 2.3. dimana penelitian ini tidak berfokus pada perhitungan prediksi tetapi berfokus pada analisis hasil, dikarenakan penelitian ini menggunakan *IBM Cloud* sebagai aplikasi pihak ke tiga, sehingga tidak ada perhitungan rumus - rumus.

2.7. *IBM Cloud*

Ibm cloud adalah sebuah platform cloud yang menawarkan berbagai layanan dan produk, termasuk komputasi, penyimpanan, analitik, kecerdasan buatan, dan banyak lagi. Karena platform ini berbasis cloud untuk mengakses berbagai layanan dan produk melalui internet tanpa perlu menginstal dan merawatnya di perangkat masing - masing. Selain itu, platform ini mengintegrasikan *infrastructure as a service* (IaaS), *platform as a service* (PaaS), dan *software as a service* (SaaS).[18] Ini memberikan ekosistem yang tak tertandingi bagi perusahaan untuk membangun, menyebarkan, dan mengelola aplikasi di jaringan pusat data di seluruh dunia.[18]

Penelitian ini menggunakan *ibm cloud* dan melibatkan beberapa servis *ibm cloud* seperti *watson machine learning with (auto ai)*, *watson studio cloud*, *object storage*, dan *watsonx*.

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.

2.8. Dataset

Dataset adalah kumpulan data terorganisir yang umumnya terkait dengan topik tertentu. Data ini bisa berupa angka, teks, gambar, atau bahkan suara. Biasanya dataset tidak cuma berisi data mentah, tapi sudah disusun dalam format tertentu, misalnya tabel. Dalam penelitian ini dataset yang digunakan adalah data tarif jalan tol.

Data tarif jalan tol yang digunakan adalah data yang diterbitkan oleh Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat di website bpjt.pu.go.id. [19][20] Selain itu ada juga data tarif jalan tol yang sudah berbentuk file .csv berasal dari kaggle.com yang ditulis oleh Muhammad Yusuf Aristyanto. [21] Namun penelitian ini hanya menggunakan data tarif tol Jabodetabek untuk diolah dan dilatih agar bisa menghasilkan *machine learning* prediksi.

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.

2.9. Penelitian Terkait

Tabel 1 Penelitian Terkait

No	Nama dan Tahun	Judul	Topik	Metode	Hasil
1	Zahedian, Sara., Nohekhan, Amir., & Farokhi Sadabadi, Kaveh. 2021.	<i>Dynamic toll prediction using historical data on toll roads: Case Study of the I-66 Inner Beltway.</i>	Prediksi Tarif Tol	<ul style="list-style-type: none"> • <i>Random Forest</i> (RF) • <i>Multilayer Perceptron</i> (MLP) • <i>Long Short-Term Memory</i> (LSTM) 	Penelitian ini membandingkan berbagai model prediksi untuk harga tol di jalan tol I-66, dengan model <i>random forest</i> berkinerja paling baik dalam memprediksi harga tol pada interval waktu yang berbeda di masa yang akan datang.
2	Harsiti, Muttaqin Zaenal, Srihartini Ela. 2021	Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet.	Prediksi Persediaan Obat Jenis Tablet	Regresi Linier Sederhana	Hasil dari penelitian ini adalah metode regresi linier sederhana dapat menentukan hasil prediksi yang akurat untuk memprediksi stok obat, dengan akurasi sebesar 98,505%.
3	Mikhael, Andreas Felix, Enri Ultach. 2022	Perbandingan Algoritma <i>Linear Regression</i> , <i>Neural Network</i> , <i>Deep</i>	Prediksi <i>Bitcoin</i>	<ul style="list-style-type: none"> • <i>Linear Regression</i> • <i>Neural Network</i> • <i>Deep Learning</i> 	Penelitian ini membandingkan berbagai model prediksi untuk bitcoin, dengan model <i>linear regression</i> yang memegang hasil prediksi terbaik.

		<i>Learning, Dan K-Nearest Neighbor (K-NN) Untuk Prediksi Bitcoin.</i>		<ul style="list-style-type: none"> • <i>K-Nearest Neighbor (K-NN)</i> 	
4	Jange Beno. 2022	Prediksi Harga Saham Bank BCA Menggunakan <i>XGBoost</i> .	Prediksi Saham	<i>XGBoost</i>	Hasil dari penelitian ini menunjukkan akurasi prediksi yang cukup baik dengan MAPE sebesar 4.01% dengan penyetelan hiper parameter; prediksi sedikit kurang baik pada bulan Maret tahun 2020 karena ada kasus pandemi Covid-19.
5	J.Avanijaa, Gurram Sunithab, K.Reddy Madhavic, Padmavathi Korad, dan R.Hitesh Sai Vittal. 2021	<i>Prediction of House Price Using XGBoost Regression Algorithm</i>	Prediksi Harga Rumah	<i>XGBoost</i>	Penelitian ini menunjukkan hasil yang menjanjikan dalam mengurangi tingkat kesalahan pengujian, yang dapat membantu pelanggan dalam mengambil keputusan yang tepat dalam transaksi <i>real estate</i> . Selain itu, penelitian ini mempertimbangkan berbagai faktor seperti area lokal, pendapatan, kekayaan, dan teknik pembelajaran mesin

					yang canggih dalam memperkirakan harga rumah.
--	--	--	--	--	---

Penelitian ini merupakan penelitian yang berfokus pada analisa hasil, maka dari itu peneliti mewajibkan jurnal pembanding atau jurnal referensi seperti hukum - hukum yang mengatur jalan tol, rumus - rumus prediksi, hingga kutipan - kutipan yang sebagai acuan dalam penelitian ini.

Penelitian dengan jurnal *Dynamic toll prediction using historical data on toll roads: Case Study of the I-66 Inner Beltway* atau jurnal pertama terdapat perbedaan. Perbedaan tersebut adalah penelitian ini menggunakan metode regresi linier dengan algoritma *extreme gradient boosting regressor* atau biasa disebut *xgboost regressor* sedangkan jurnal pertama menggunakan tiga metode, yaitu : *Random Forest (RF)*, *Multilayer Perceptron (MLP)*, *Long Short-Term Memory (LSTM)*. Meskipun pada percobaan kedua algoritma yang digunakan sama, selain itu penelitian ini memprediksi tarif tol menggunakan variabel ruas jalan tol, panjang jalan, dan golongan kendaraan; sedangkan jurnal pertama memprediksi harga tol dan perbedaan waktu tempuh di jalan tol I-66 dan rute alternatifnya. Oleh karena itu penelitian ini menggunakan referensi jurnal pertama ini karena pada percobaan kedua algoritma yang digunakan sama.

Untuk jurnal Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet atau jurnal kedua ini sama - sama menggunakan metode regresi linier tetapi topik yang diprediksi berbeda penelitian ini memprediksi tarif jalan tol, sedangkan jurnal kedua memprediksi persediaan obat.

Lalu perbedaan penelitian ini dengan jurnal Perbandingan Algoritma *Linear Regression, Neural Network, Deep Learning, Dan K-Nearest Neighbor (K-NN)* Untuk Prediksi *Bitcoin* atau jurnal ketiga adalah penelitian ini memprediksi tarif tol sedangkan jurnal ketiga memprediksi *bitcoin*. Di jurnal ketiga membandingkan berbagai model prediksi untuk bitcoin, dengan model *linear regression* yang memegang hasil prediksi terbaik. Oleh karena itu penelitian ini menggunakan referensi jurnal ketiga ini karena diantar model - model prediksi model *linear regression* yang memegang hasil prediksi terbaik, yang mana sama model ini dipakai dalam penelitian ini.

Kemudian jurnal Prediksi Harga Saham Bank BCA Menggunakan *XGBoost* atau jurnal keempat yang ditulis oleh Beno Jange adalah perbedaan topik dimana penelitian ini memprediksi tarif tol sedangkan jurnal keempat memprediksi saham. Juga jika penelitian ini menggunakan model regresi sedangkan jurnal ketiga menggunakan model deret waktu. Tetapi algoritma yang digunakan sama – sama menggunakan *xgboost*, jadi pada jurnal keempat ini dijadikan referensi dan mengutip dari jurnal keempat.

Dan yang terakhir jurnal *Prediction of House Price Using XGBoost Regression Algorithm* atau jurnal kelima yang ditulis oleh J.Avanijaa, Gurram Sunithab, K.Reddy Madhavic, Padmavathi Korad, dan R.Hitesh Sai Vittal dimana jurnal kelima ini bukan perbedaan melainkan sebuah jurnal acuan untuk algoritma *xgboost* teori - teori *xgboost regressor* dikutip dari jurnal kelima ini. Namun masih terdapat perbedaan yaitu topik yang dibahas jika penelitian ini memprediksi tarif jalan tol golongan I sedangkan jurnal kelima ini memprediksi harga rumah.

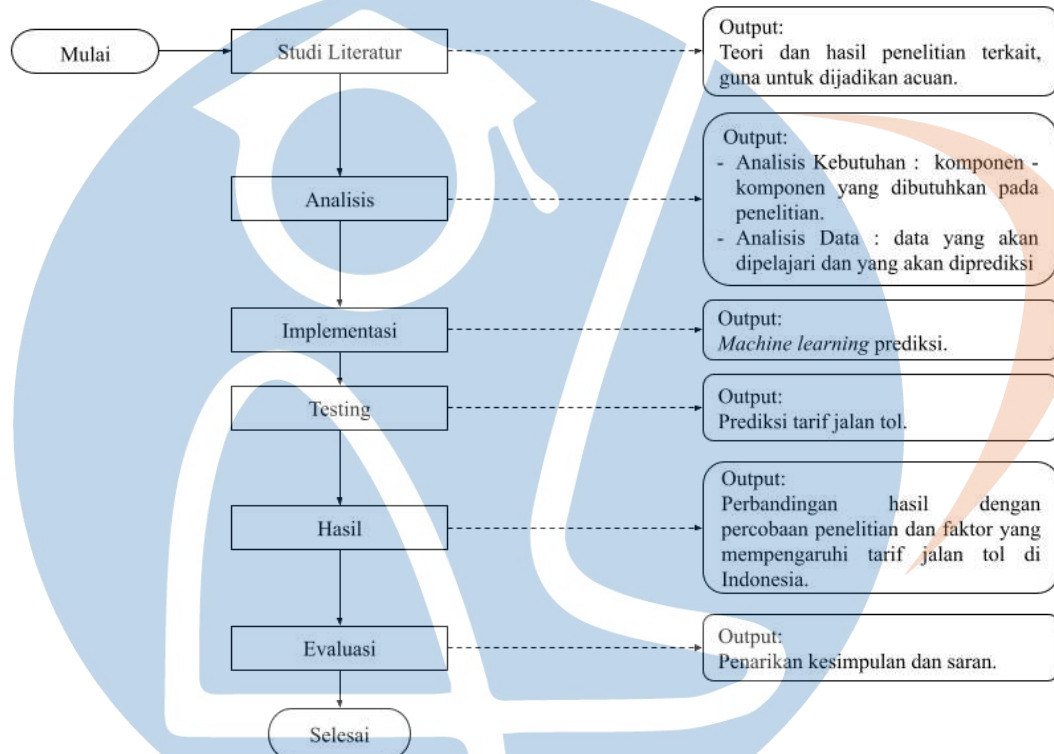


STT - NF

BAB III METODOLOGI PENELITIAN

3.1. Tahapan Penelitian

Tahapan penelitian yang dilakukan untuk memprediksi tarif tol golongan I dapat dilihat pada gambar berikut.



Gambar 2 Tahapan Penelitian

Penelitian ini dilakukan dengan tahapan – tahapan penelitian yang meliputi studi literatur, analisis kebutuhan, analisis data, implementasi, testing, hasil, dan evaluasi.

3.1.1. 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

3.1.2. 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*. Analisa dilakukan dengan cara membuat servis - servis yang ada di *ibm cloud* menjadi resource. 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.

3.1.3. 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.

Seperti yang dijelaskan pada sub bab 2.8., 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.

3.1.4. Implementasi

Tahap selanjutnya implementasi dengan menggunakan servis - servis yang dijelaskan pada sub bab 3.1.2. 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 bab iv alur pengerjaan.

3.1.5. Testing

Tahap selanjutnya adalah testing dari *machine learning* prediksi yang sudah dibuat. Testing ini dilakukan akan memprediksi Tarif Golongan I dengan variabel 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.

3.1.6. 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 dari percobaan - percobaan yang dilakukan. 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.

3.1.7. 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.

3.2. Jenis Penelitian

Penelitian eksperimen bertujuan untuk menyelidiki hubungan sebab - akibat antara variabel - variabel. Ini dicapai dengan mengontrol dan mengubah variabel independen serta mengamati bagaimana variabel dependen berdampak pada mereka. Menurut Suharsimi Arikunto, eksperimen berarti suatu cara untuk mencari hubungan sebab akibat (hubungan kausal) antara dua faktor yang sengaja ditimbulkan oleh peneliti dengan mengeliminasi atau mengurangi atau menyisihkan faktor - faktor lain yang mengganggu.[22, hlm. 9] Menurut Sugiyono, penelitian eksperimen adalah salah satu metode kuantitatif yang digunakan ketika peneliti

ingin melakukan percobaan untuk mencari pengaruh variabel independen atau perlakuan terhadap variabel dependen atau hasil dalam kondisi yang terkendali.[23, hlm. 110] Tujuan penelitian eksperimen selain menyelidiki hubungan sebab - akibat antara variabel - variabel adalah.

- **Menguji hipotesis:** Hipotesis adalah pernyataan yang diajukan oleh peneliti tentang hubungan sebab-akibat antara variabel bebas dan variabel terikat.
- **Menemukan hubungan kausal:** Penelitian eksperimen bertujuan untuk menemukan hubungan kausal, yaitu hubungan sebab-akibat antara variabel bebas dan variabel terikat.
- **Membuat generalisasi:** Hasil penelitian eksperimen yang baik dapat digeneralisasikan ke populasi yang lebih luas.

Oleh karena itu, penulis memilih jenis penelitian ini karena penelitian ini dilakukan secara sistematis dengan mengumpulkan, mengolah, menganalisis, dan menyajikan data. Ini sesuai dengan cara melakukan kegiatan penelitian eksperimen. Dan data yang digunakan dalam penelitian ini berupa variabel - variabel yang saling berhubungan dan variabel apa yang akan mempengaruhi hasil prediksi, dan seberapa besar variabel tersebut mempengaruhi hasil prediksi. Ini juga sesuai dengan pengertian penelitian eksperimen.

3.3. Metode Analisis Data

Penulis menggunakan penelitian kuantitatif dikarenakan penelitian kuantitatif menggunakan pengukuran objektif dan analisis statistik terhadap sampel data yang diperoleh melalui tes, kuesioner, jejak pendapat, atau instrumen penelitian lainnya. Tujuan dari metode ini adalah untuk membuktikan atau menguji hipotesis (dugaan sementara) yang diajukan dalam penelitian. Menurut Sugiyono, penelitian kuantitatif adalah metode penelitian yang didasarkan pada filsafat positivisme, sebagai metode ilmiah atau scientific karena memenuhi kaidah ilmiah secara konkret atau empiris, objektif, terukur, rasional, dan sistematis.[23, hlm. 16] Dan penelitian kuantitatif didasarkan pada filsafat positivisme dan memerlukan dua kriteria penting: eksplanatori dan prediktif. Kriteria eksplanatori berarti bahwa

penelitian harus dapat menjelaskan bagaimana dua fenomena atau lebih berhubungan satu sama lain dalam bentuk hubungan, perbedaan, atau pengaruh. Kriteria prediktif berarti bahwa penelitian harus dapat menjelaskan bagaimana sampel penelitian dibandingkan dengan populasinya. Prediktif berarti bahwa hasil penelitian harus memiliki kemampuan untuk memprediksi fenomena yang akan terjadi.

Berdasarkan kriteria - kriteria yang dijelaskan sebelumnya, metode penelitian kuantitatif sangat cocok pada penelitian ini, karena kriteria - kriteria tersebut merupakan fokus dan tujuan penulis pada penelitian ini. Dimana penelitian ini memprediksi lalu membandingkan data sampel atau data hasil pengujian dengan data populasi atau data asli, dan variabel apa yang paling mempengaruhi hasil pengujian.

3.4. Metode Pengumpulan Data

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 kuantitatif 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. Seperti yang sudah dijelaskan pada sub bab 2.8, 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. [19][20] Selain itu ada juga data tarif jalan tol yang sudah berbentuk file .CSV berasal dari kaggle.com yang ditulis oleh Muhammad Yusuf Aristyanto. [21]

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 itupun per daerah dan tidak ada *latitude* dan *longitude*, Jadi mengandalkan google maps untuk informasi *latitude* dan *longitude*.

3.5. Metode Pengujian

Pada penelitian ini pengujian dilakukan uji akurasi pada model menggunakan metrik, dan pada penelitian kali ini metrik yang digunakan adalah *root mean squared error* (RMSE), dimana 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 dimana 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.

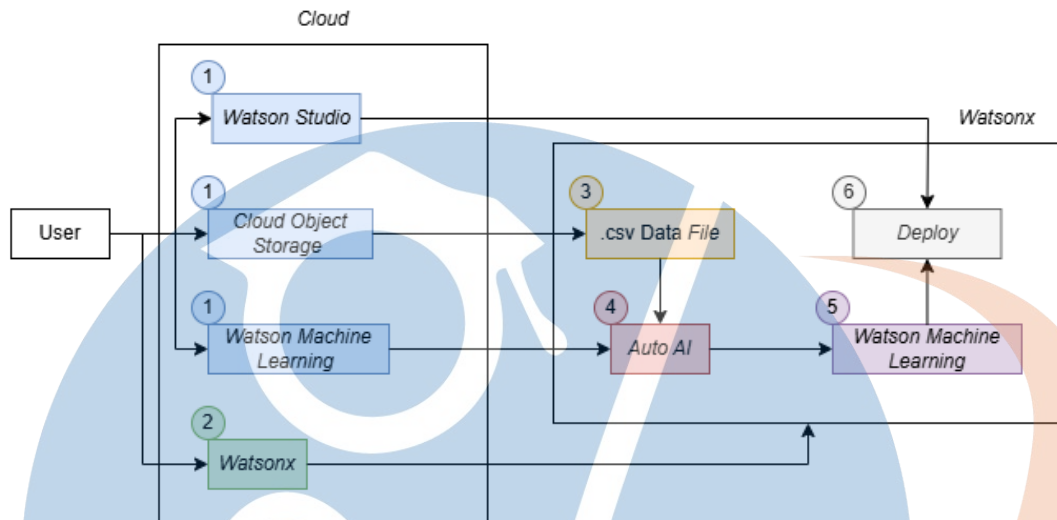
4. ***Mean Squared Log Error*** adalah metrik yang digunakan dalam masalah regresi untuk mengukur kesalahan prediksi dalam bentuk logaritmik.
5. ***Mean Absolute Error*** adalah metrik statistik yang digunakan untuk mengukur rata-rata kesalahan absolut antara nilai prediksi model dan nilai aktual.
6. ***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.

STT - NF

BAB IV IMPLEMENTASI DAN HASIL

4.1. Alur Pengerjaan



Gambar 3 Alur Pengerjaan

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*.

4.2. Implementasi

Require service:

- *Watson Machine Learning (Auto AI)*
- *Watson Studio*
- *Cloud Object Storage*
- *Watsonx*

Data Format : file .csv

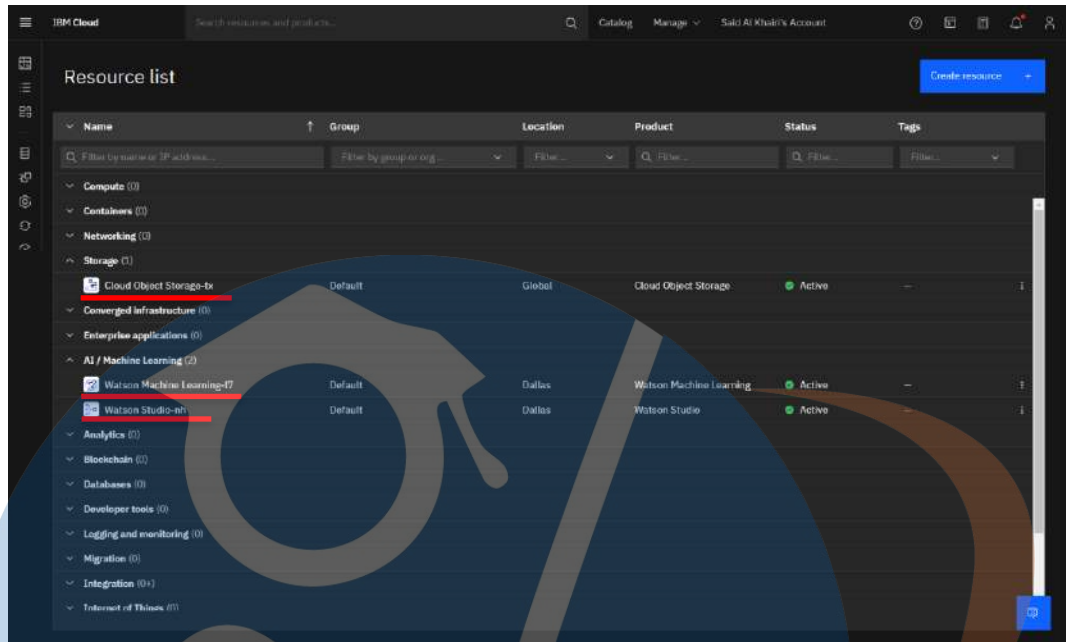
Disini penulis menggunakan data Jabodetabek Toll Road Rates (Gol I).csv seperti pada gambar berikut.

	A	B	C	D	E	F	G
	Ruas Jalan Tol	Asal Perjalanan	Tujuan Perjalanan	Panjang (Km)	Latitude	Longitude	Alamat
1	Akses Tanjung Priok	Akses Tanjung Priok Seksi E-1	Akses Tanjung Priok Seksi E-2	3,40	-6,1195274	106,8932664	Jl. Raya Sulawesi No.6, RT.6/RW.13, Raw...
2	Akses Tanjung Priok	Akses Tanjung Priok Seksi E-2	Akses Tanjung Priok Seksi NG	2,74	-6,1084388	106,9025016	Jl. Jampoe 1, Koja, Kec. Koja, Jkt Utara, D...
3	Akses Tanjung Priok	Akses Tanjung Priok Seksi E-2A	Akses Tanjung Priok Seksi E-2	1,92	-6,1040771	106,8999844	VVWX+9X9, Jl. Timor Raya, RW.1, Koja, K...
4	Akses Tanjung Priok	Akses Tanjung Priok Seksi NS	Akses Tanjung Priok Seksi E-2A	2,24	-6,1083657	106,912905	VWR6+M X5, Lingkar Luar Timur, RT.1/RW...
5	Akses Tanjung Priok	Akses Tanjung Priok Seksi NS (Rorotan Kebon Bawang)	Akses Tanjung Priok Seksi E-2A	1,10	-6,1179397	106,9250087	VWJG+R6C, RW.3, East Semper, Cilincing...
6	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,2871125	106,8782589	PV7H+582, Jl. Jkt - Cibodok, RT.17/RW.1...
7	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,2875338	106,8776752	PV6H+X3M, RT.4/RW.1, Dukuh, Kramat J...
8	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,2960158	106,8783167	Dukuh, Kramat Jati, East Jakarta City, J...
9	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3024932	106,8830419	Jakarta Timur, MVXM+263, RT.12/RW.4, [...]
10	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,2906532	106,8756233	Jl. Dukuh III, Dukuh, Kec. Kramat Jati, Kota...
11	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3101479	106,8847408	MVQM+WVX, RT.11/RW.5, Rambutan, Cite...
12	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3560872	106,8797992	Cibubur, Ciracas, East Jakarta City, Jakt...
13	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3552993	106,8942258	JVMV+YMQ, RT.8/RW.7, Cibubur, Ciracas...
14	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3549369	106,8428272	Cimanggis, Depok City, West Java, Indone...
15	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3827938	106,8957389	JV8W+V7Q, Harjamukti, Cimanggis, Depo...
16	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3866476	106,8964882	JV7W+8HX, Harjamukti, Cimanggis, Depo...
17	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3852023	106,8955138	Jl. Tol Jagorawi No.22, RT.9/RW.9, Harja...
18	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3876722	106,898188	JV6X+WJF, Jl. Tol Jagorawi, Harjamukti, [...]
19	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,4612768	106,8693438	GV00+FGJ, Jl. Raya Gn. Putri, Gn. Putri, K...
20	Jakarta - Bogor - Ciawi	Bogor	Ciawi	23,00	-6,4572121	106,8840701	GVVM+4J8, Karanganyar, Gunung Putri, Bo...
21	Jakarta - Bogor - Ciawi	Bogor	Ciawi	23,00	-6,4824650	106,8744562	GV9F+2Q8, Puspasari, Citeureup, Bogor R...

Gambar 4 Dataset Jabodetabek Toll Road Rates (Gol I)

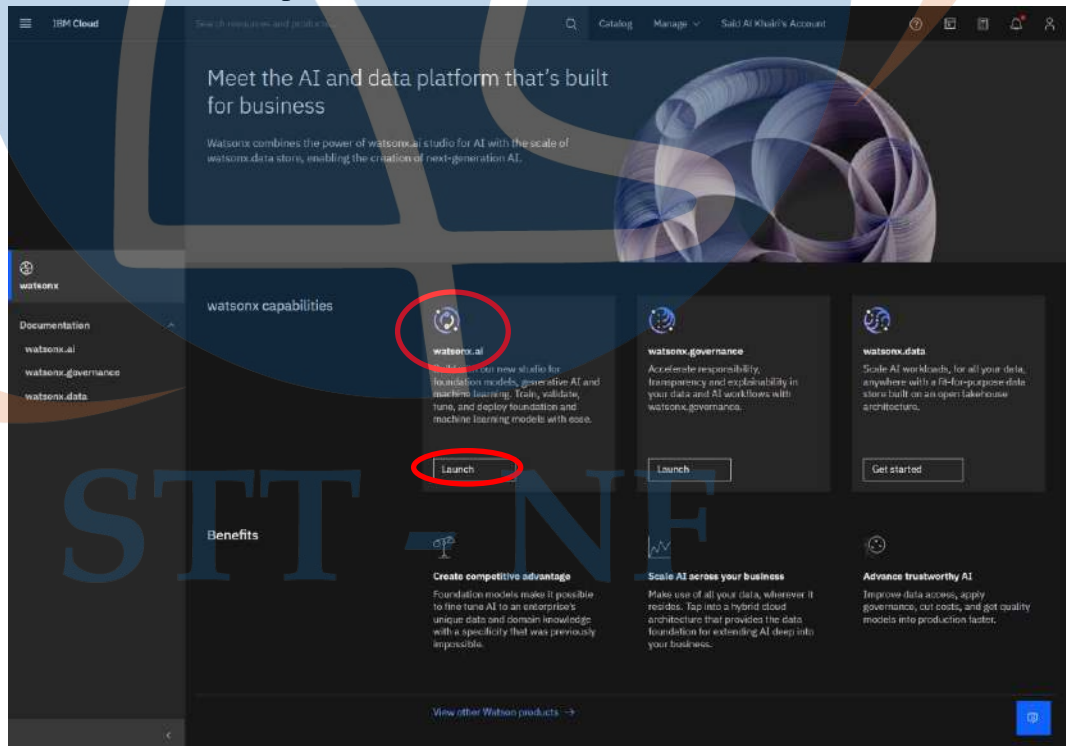
STT - NF

1. Menambahkan servis - servis yang dibutuhkan yang nanti akan tampil di *resource list* seperti pada gambar berikut.



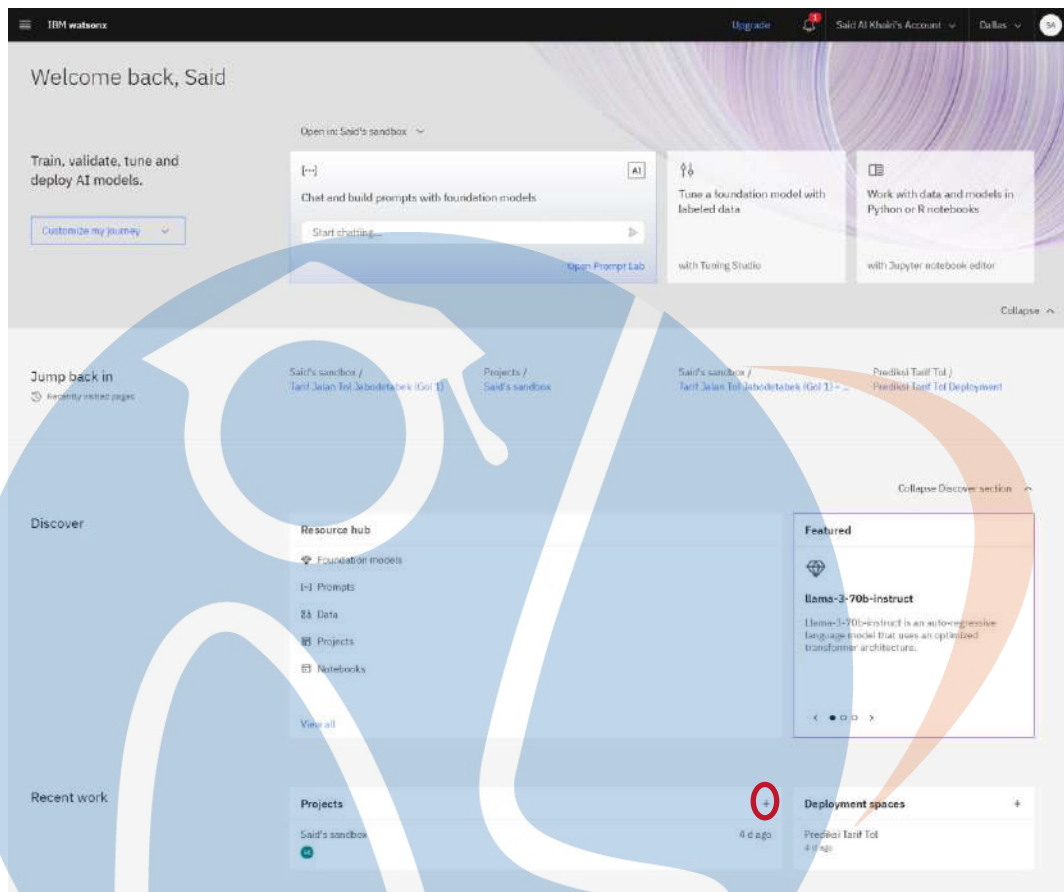
Gambar 5 Resource List

2. Jalankan *watsonx.ai* pada servis *watsonx*.



Gambar 6 Watsonx

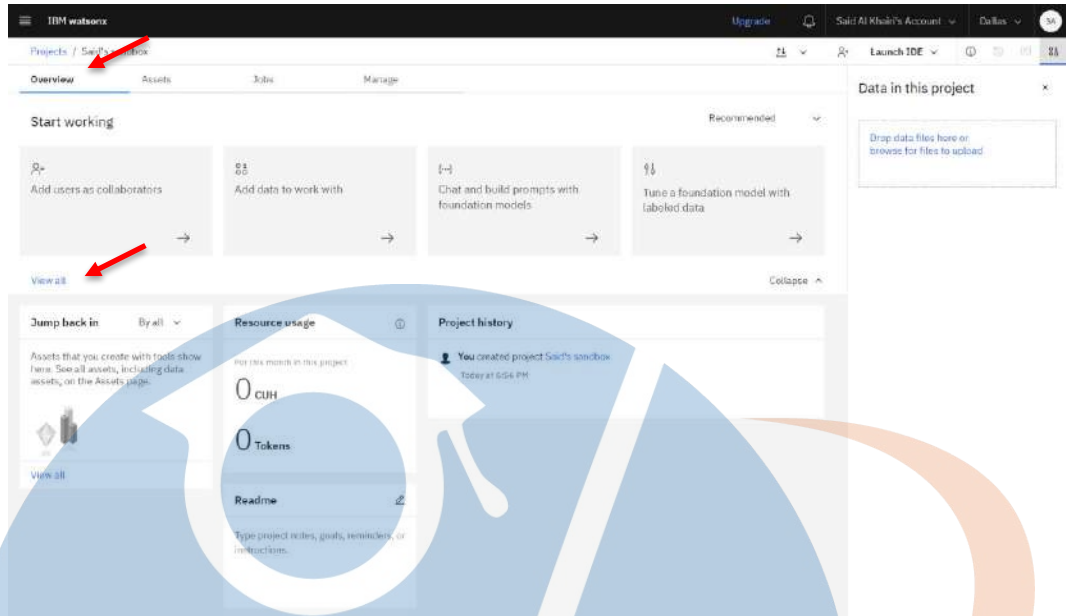
3. Berikut adalah gambar halaman *home service watsonx*, untuk memulai proyek bisa pilih tanda “+”.



Gambar 7 Home Service Watsonx

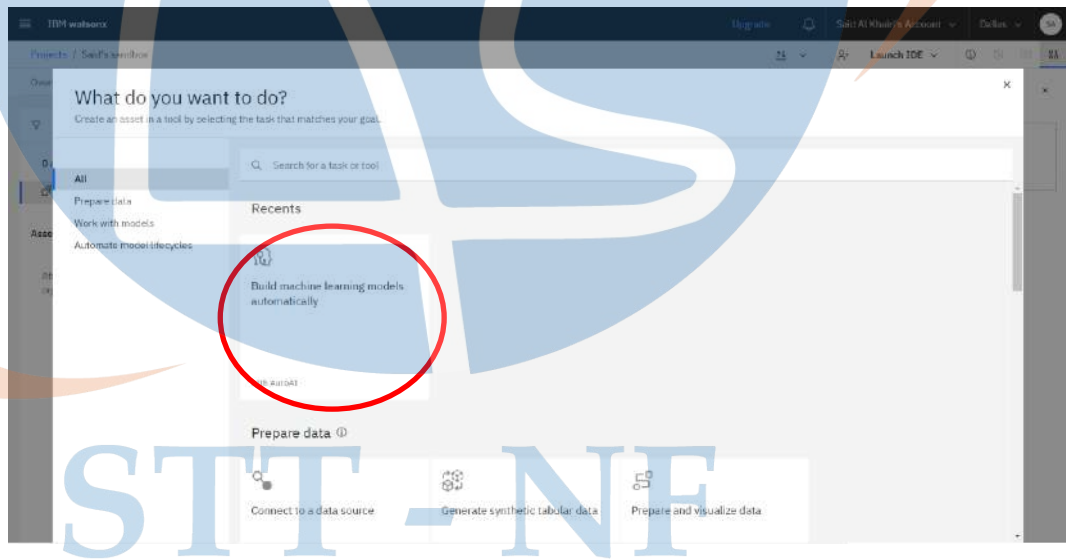
STT - NF

4. Pada halaman projects, pada menu overview, pilih view all.



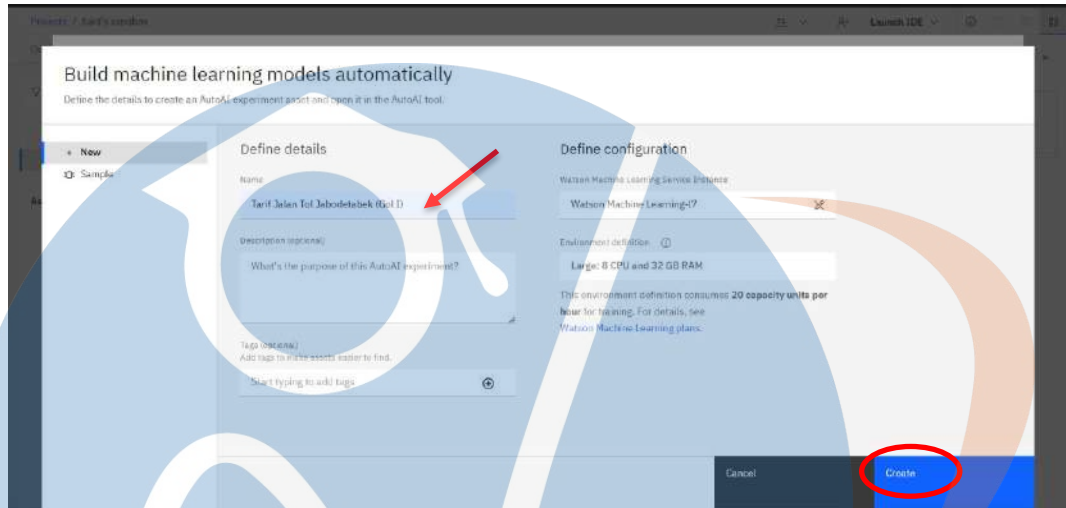
Gambar 8 Projects

5. Telusuri pada search bar “build machine learning models automatically”.



Gambar 9 Start Projects

6. Tentukan nama dan deskripsi (opsional) untuk project yang dilakukan. Pastikan *Watsonx Machine Learning Service Instance* sudah terkait dengan *Watsonx Machine Learning*. Ini dilakukan untuk mengaitkan *Watsonx Machine Learning Server* dengan *project* yang dilakukan, lalu *create*.



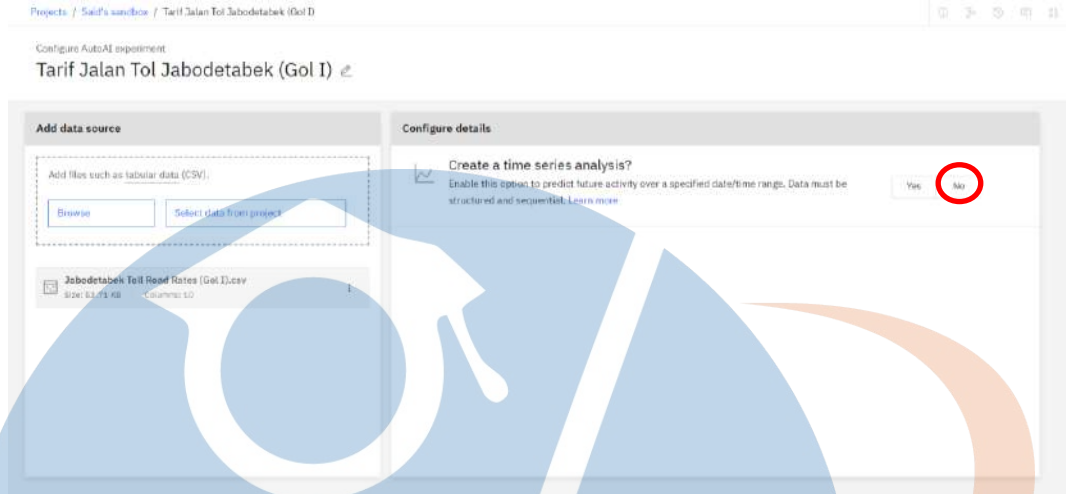
Gambar 10 Define Projects

7. Pada halaman input file. Pilih *Browse* dan masukkan dataset yang sudah disiapkan, dan pastikan file tersebut berformat *.csv*.



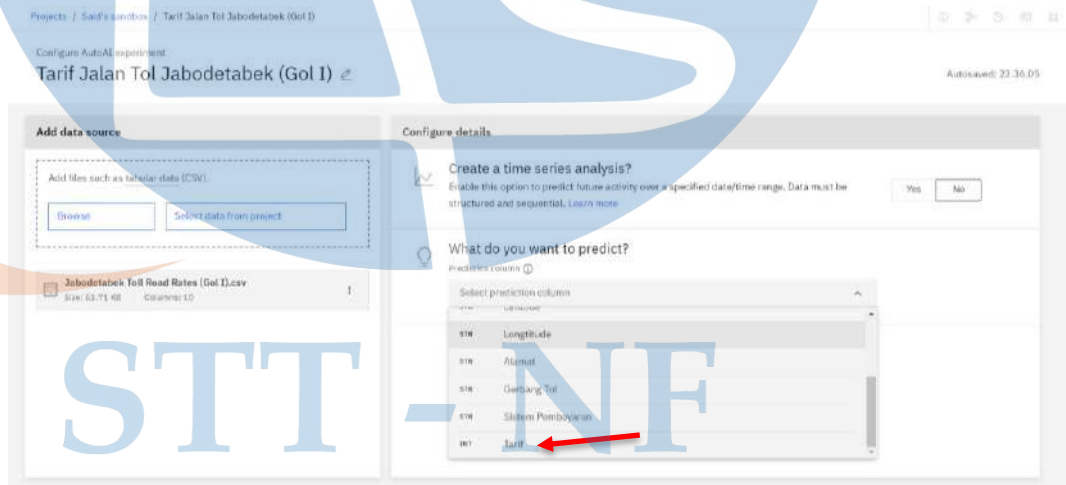
Gambar 11 Upload Dataset

8. Pada bagian *Configure details* pilih *no* untuk opsi *create a time series analysis*. Penulis memilih *no* karena penelitian ini bukan memprediksi secara historis.



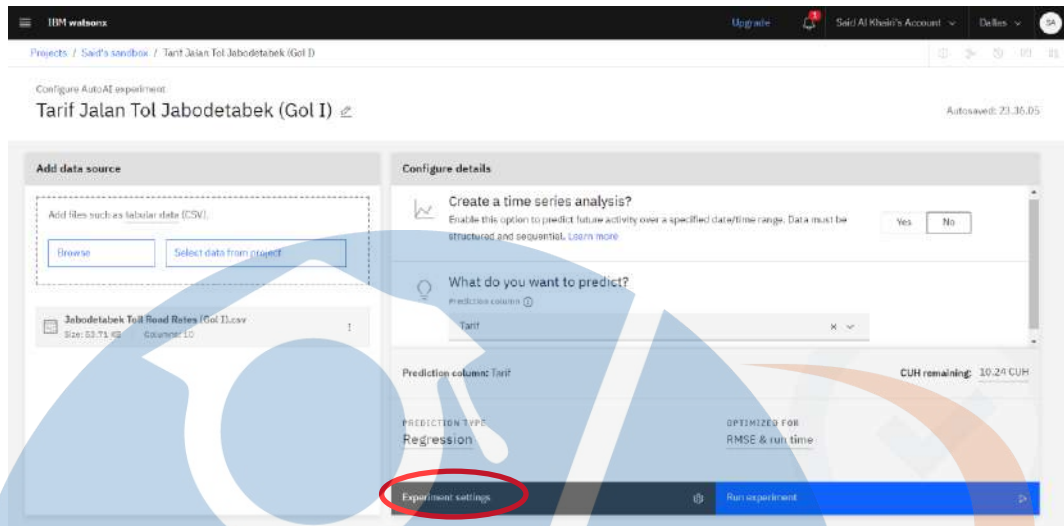
Gambar 12 Konfigurasi Model *Machine Learning*

9. Tentukan *predict column* sesuai dengan kategori yang akan digunakan dalam prediksi. Dalam penelitian ini dipilih Tarif untuk diprediksi, karena kolom Tarif merupakan kolom yang akan digunakan dalam prediksi.



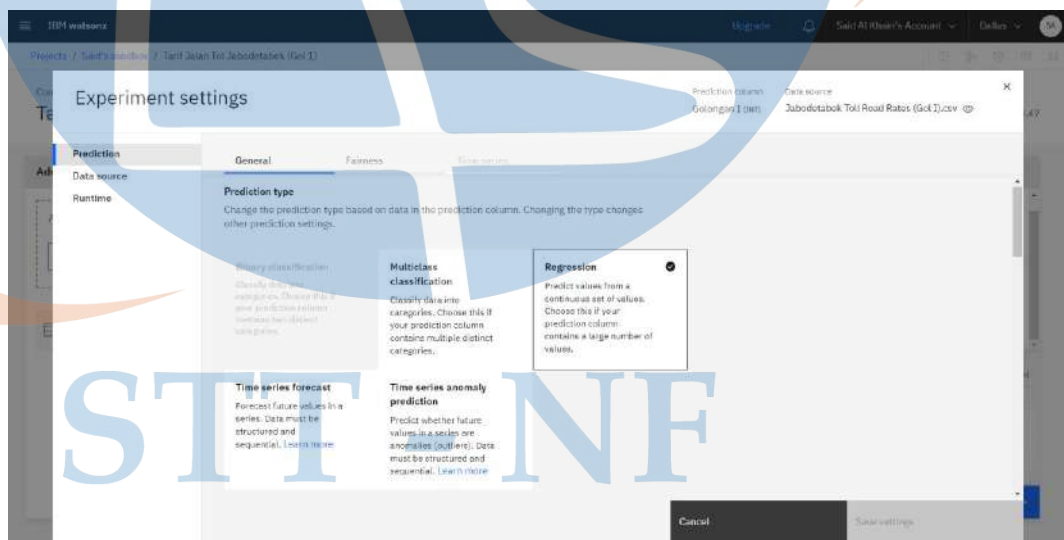
Gambar 13 Konfigurasi Model *Machine Learning*

10. Halaman akan memunculkan tipe prediksi. Pilih bagian *experiment settings* untuk pengaturan lainnya.



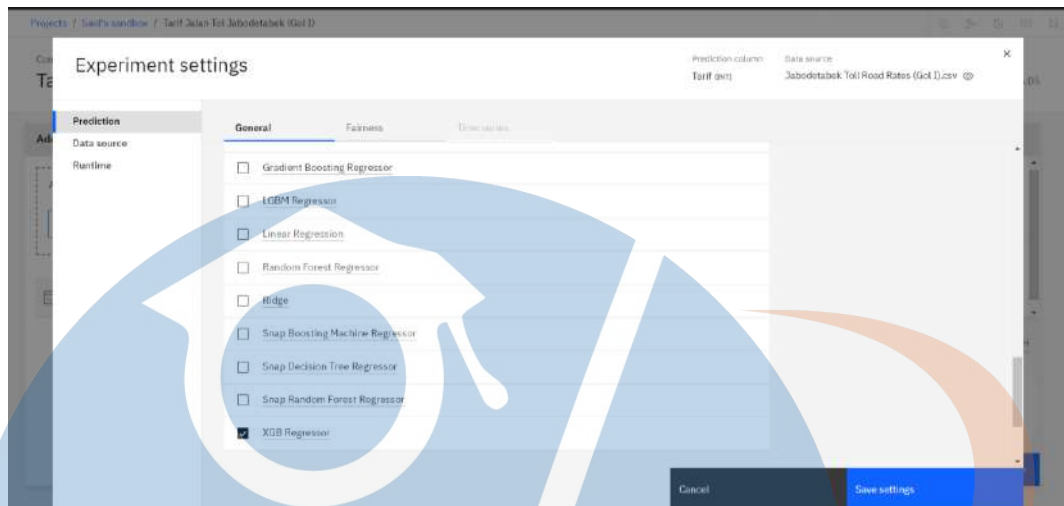
Gambar 14 Konfigurasi Model *Machine Learning*

11. Pada bagian *experiment setting*, pengguna bisa mengatur sendiri *prediction type*, algoritma yang dimasukkan, dan algoritma yang dipakai. Pilih *regression* pada bagian *prediction type* berdasarkan rekomendasi *watsonx*.



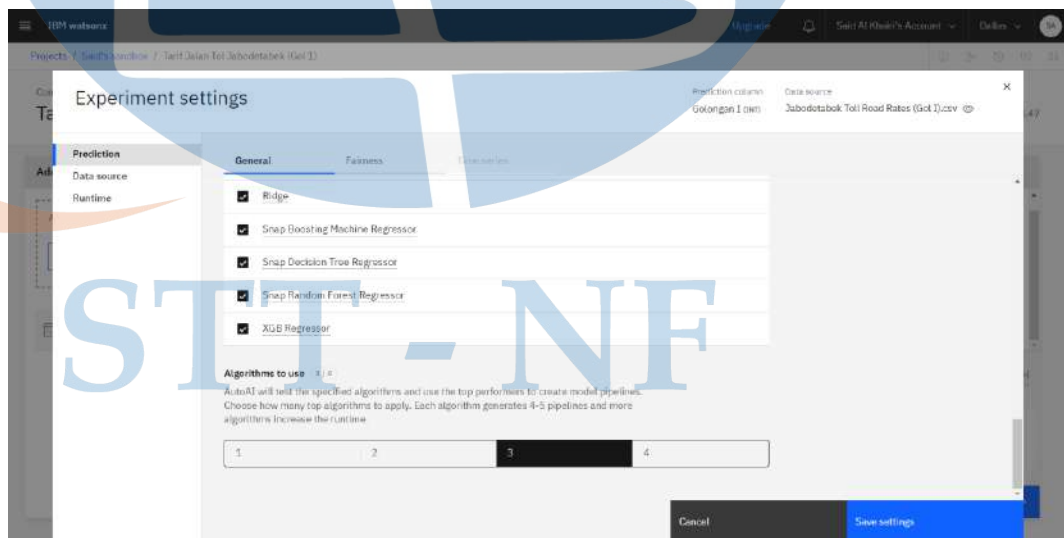
Gambar 15 *Prediction Type*

12. Pada bagian *algorithms to include* pilih algoritma yang sesuai dengan project. Dalam penelitian ini algoritma yang digunakan adalah *xgb regressor*.



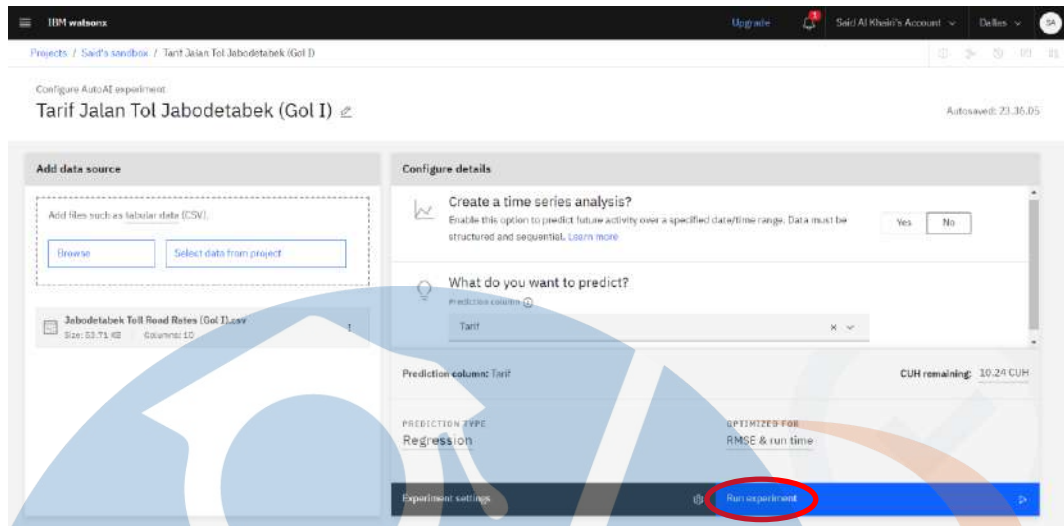
Gambar 16 Algorithms To Include

13. Pada bagian *algorithms to use* pilih algoritma yang digunakan dengan memilih angka 1-4. Semakin besar angka menandakan jumlah berapa banyak hasil algoritma yang dibuat dan hal tersebut semakin banyak pula waktu yang dibutuhkan untuk prediksi. Simpan settingan dengan klik *save settings*.



Gambar 17 Algorithms To Use

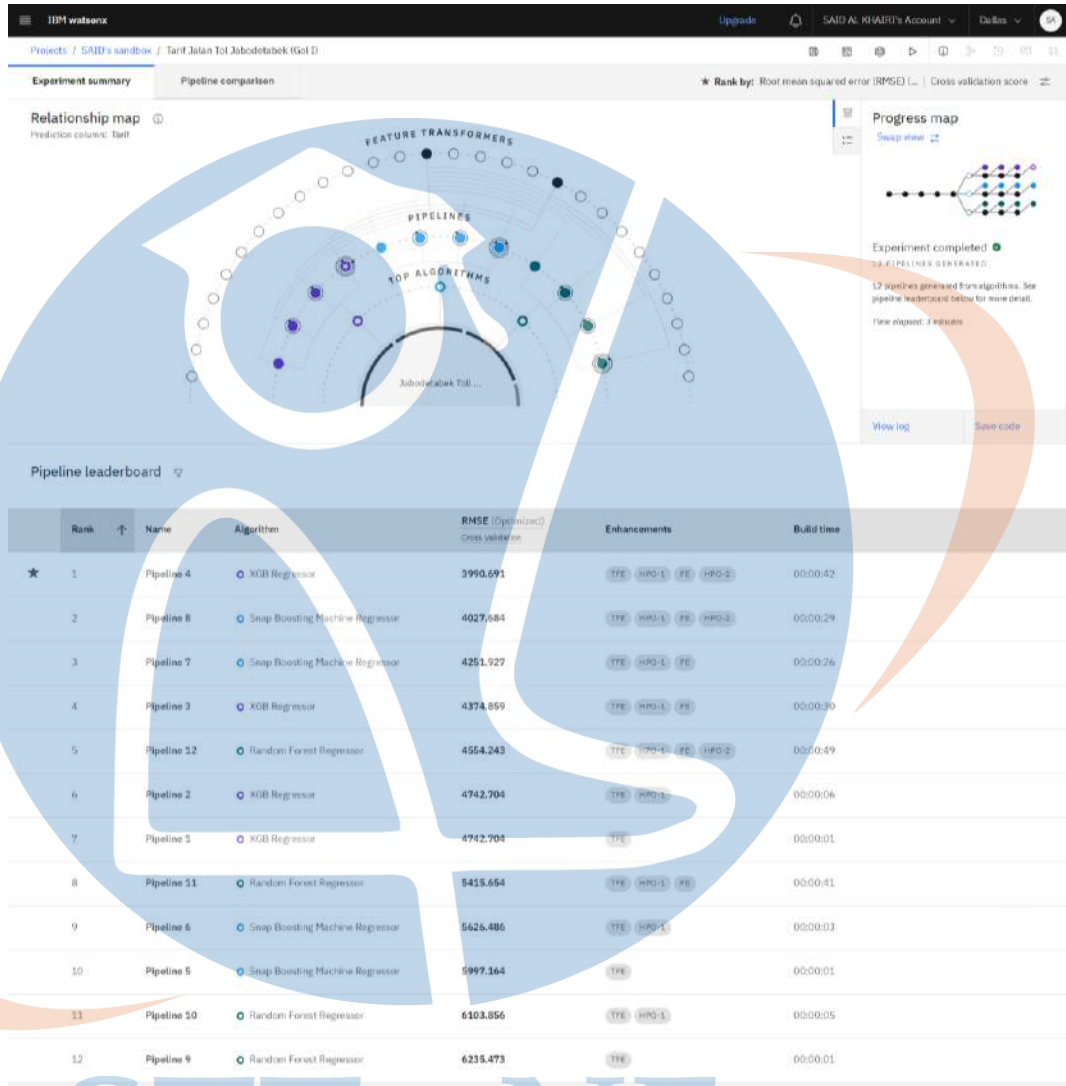
14. Jalankan eksperimen dengan pilih *run experiment*.



Gambar 18 Proses Run Experiment

STT - NF

15. Tunggu hingga selesai model dilatih menggunakan dataset nantinya akan muncul infografis dan *pipeline* dengan tingkat akurasi yang berbeda - beda. Pilih yang menempati peringkat 1 untuk melihat detailnya, dalam kasus ini *pipeline* 8 menempati peringkat pertama.



Gambar 19 Hasil *Experiment*

16. Berikut merupakan isi dari *pipeline 4 (Top Performance)*. Ada menu *model evaluation* untuk mereview dan menganalisa pada *model evaluation* dan ada juga menu *feature summary* untuk mereview dan menganalisa pada *feature summary*.

Measures	Holdout score	Cross validation score
Root mean squared error	2174.963	9093.691
R squared	0.709	0.702
Explained variance	0.724	0.706
Mean squared error	4726551.909	56172932.909
Mean squared log error	0.968	0.106
Mean absolute error	1236.627	2101.689
Median absolute error	482.767	749.465
Root mean squared log error	0.261	0.324

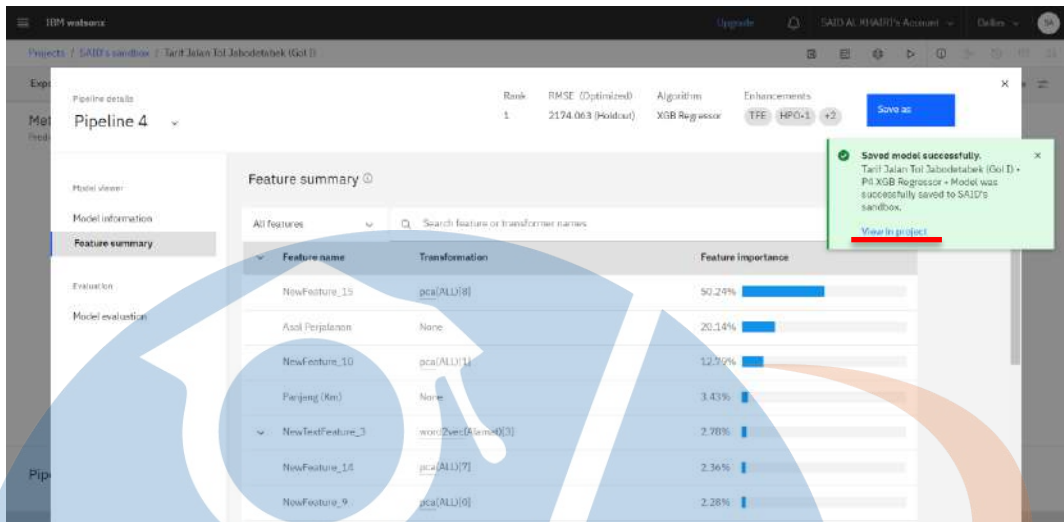
Gambar 20 Model Evaluation

17. Pada menu *feature summary* bisa dilihat dan diperhatikan pada bagian *feature importance* mengidentifikasi fitur penting yang akan mempengaruhi hasil. Untuk menyimpan model algoritma tersebut pilih *save as*.

Feature name	Transformation	Feature importance
Aksi Perjalanan	None	70.14%
Parjang (Km)	None	3.43%
Tujuan Perjalanan	None	0.89%

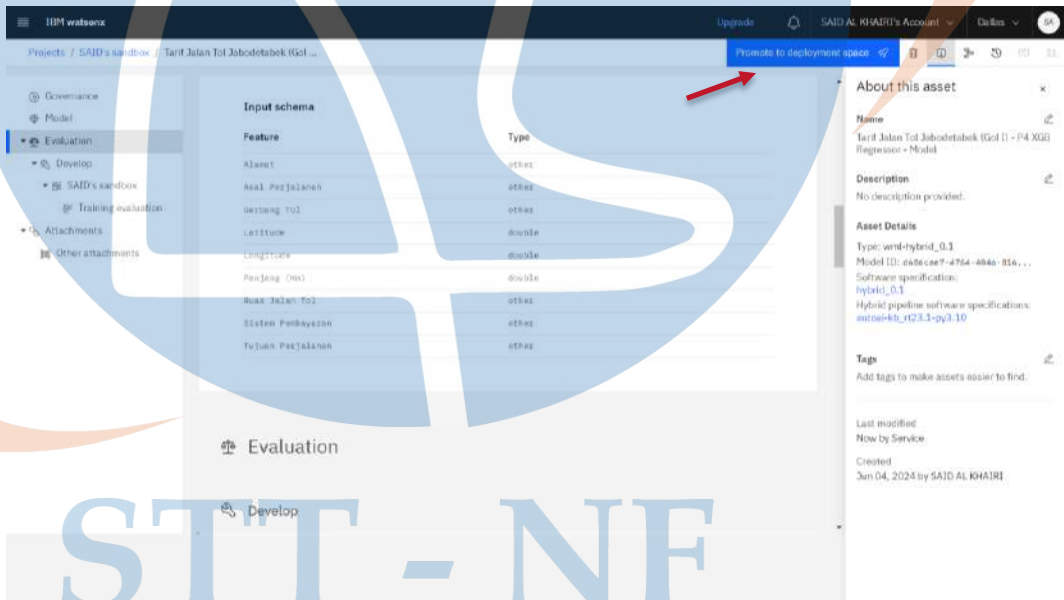
Gambar 21 Feature Summary

18. Nantinya akan ada notifikasi model sudah di simpan, pilih *view in project* pada notifikasi.



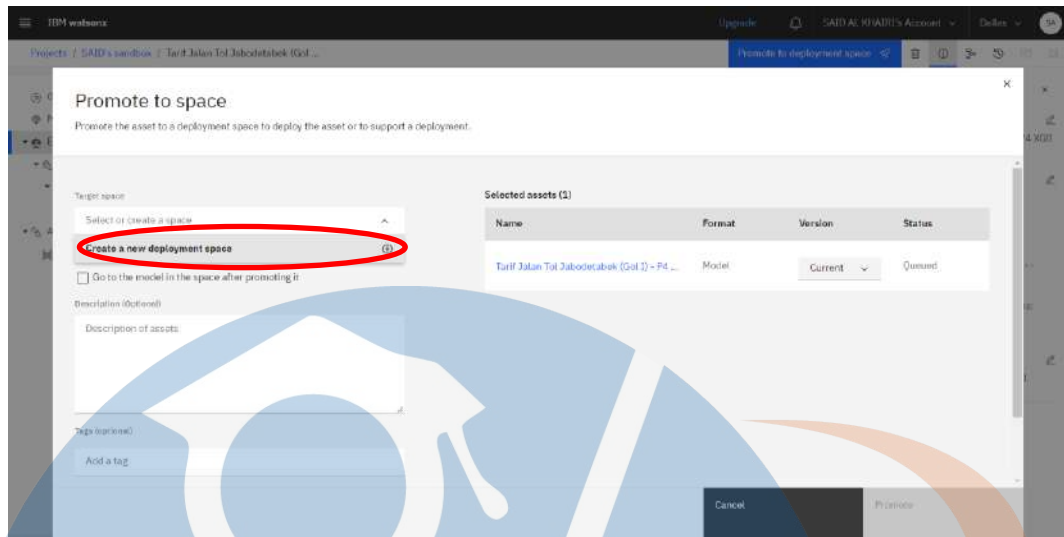
Gambar 22 Save As Untuk Deployment

19. Pilih *promote to deployment*.



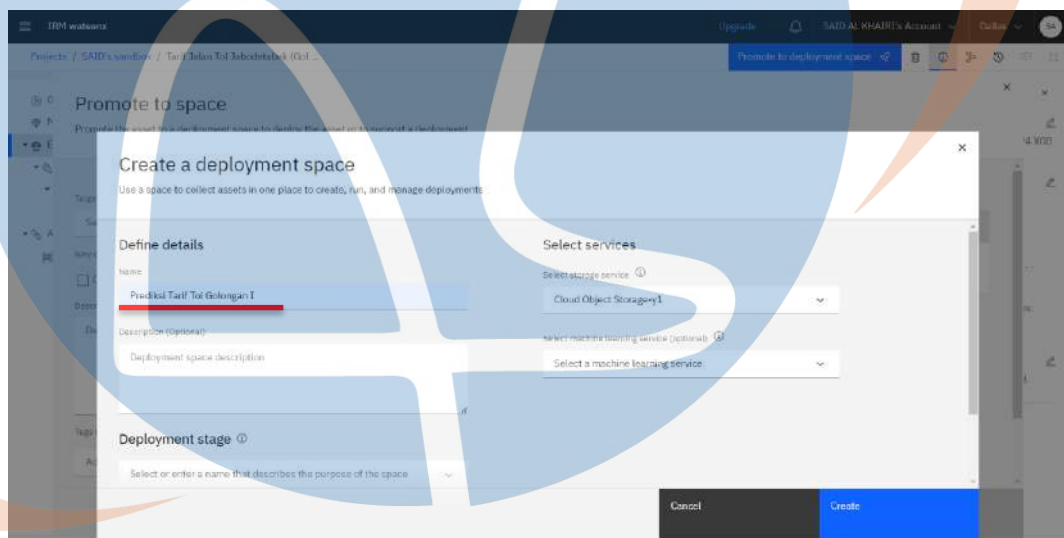
Gambar 23 Promote To Deployment

20. Tentukan target space dengan pilih *create a new deployment space*.



Gambar 24 Create New Deployment

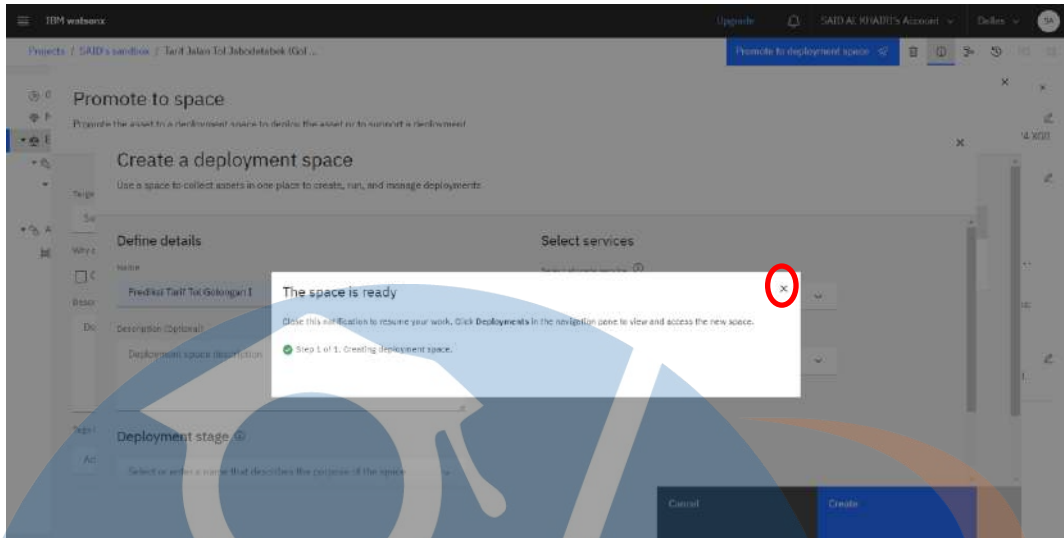
21. Tentukan nama, kaitkan dengan *machine learning service* lalu *create*.



Gambar 25 Create New Deployment

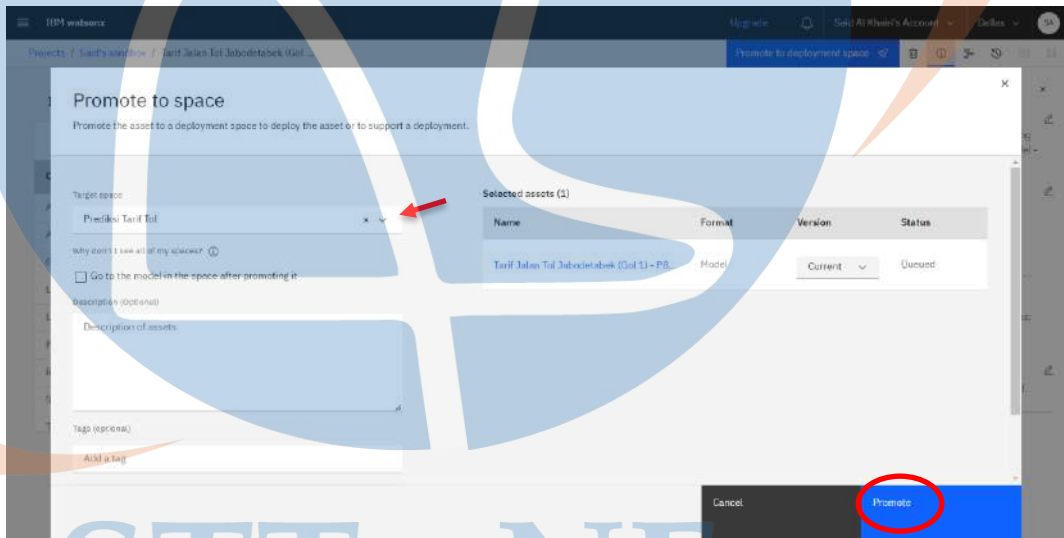
STT - NF

22. Pilih *close*.



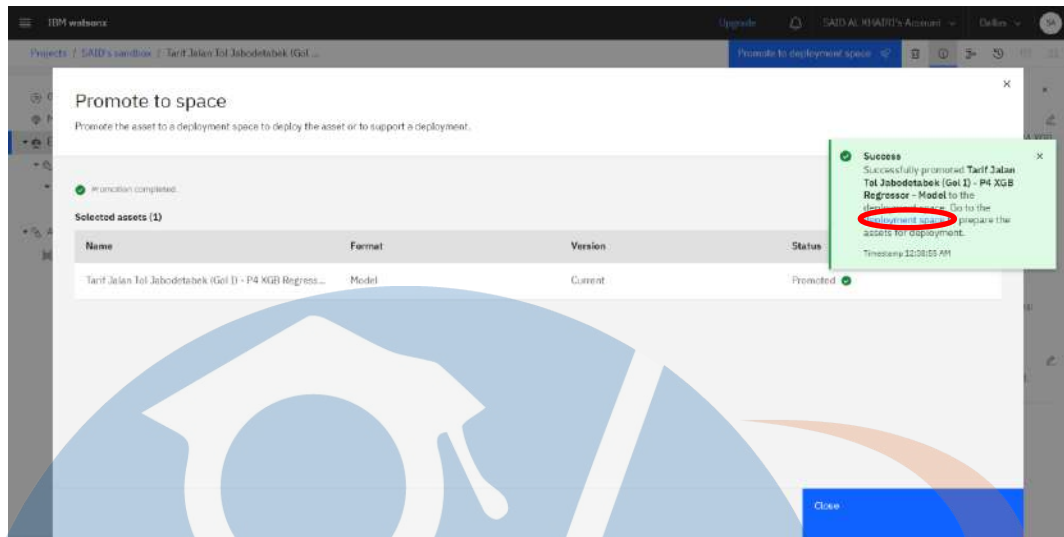
Gambar 26 Create New Deployment

23. Pada *target space* pilih target yang sebelumnya dibuat. Kemudian pilih *promote*.



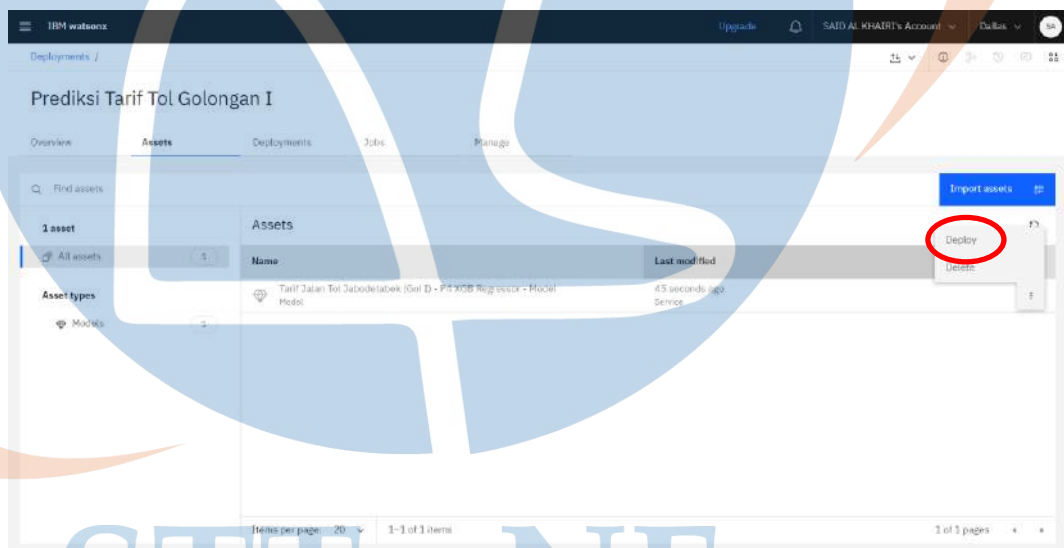
Gambar 27 Promote Deployment

24. Pada bagian notifikasi yang muncul pilih *deployment space*.



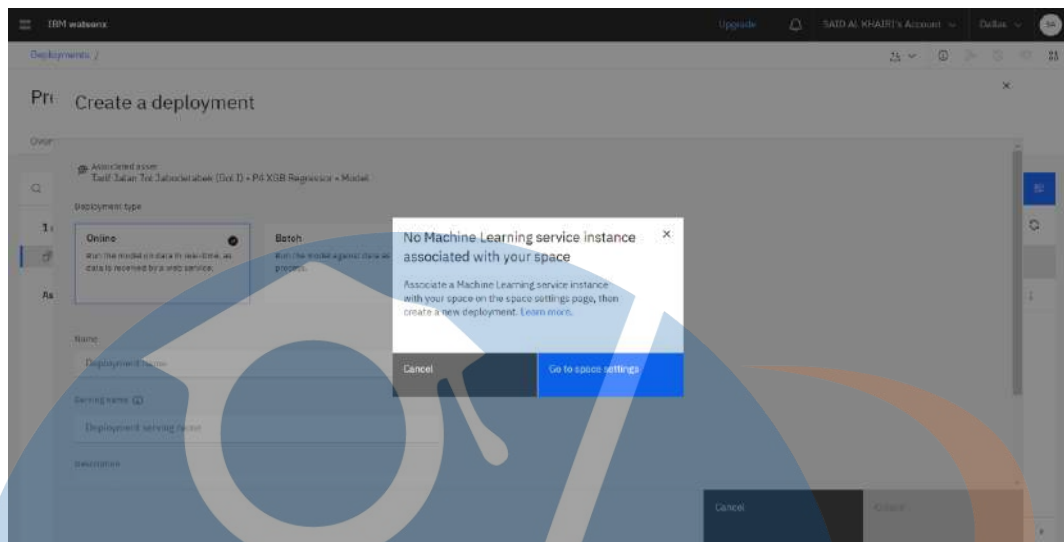
Gambar 28 Promote Deployment

25. Pilih menu *assets*, lalu pilih titik tiga pada model yang sebelumnya dibuat, kemudian pilih *deploy*.



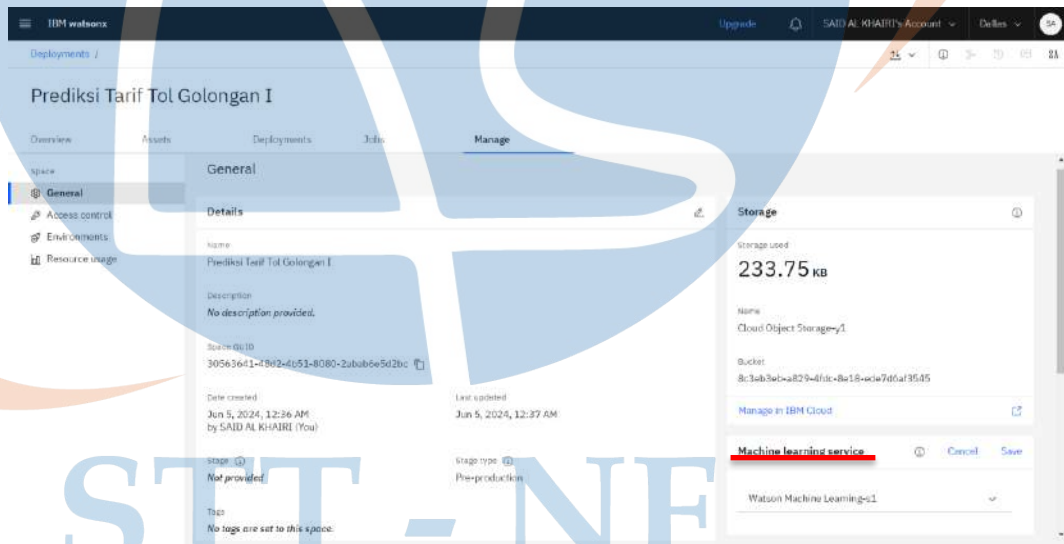
Gambar 29 Deploy Project

26. Pada bagian *create a deployment* pilih *online* untuk tipe deployment, masukkan nama (optional) lalu *create* > *go to space settings*.



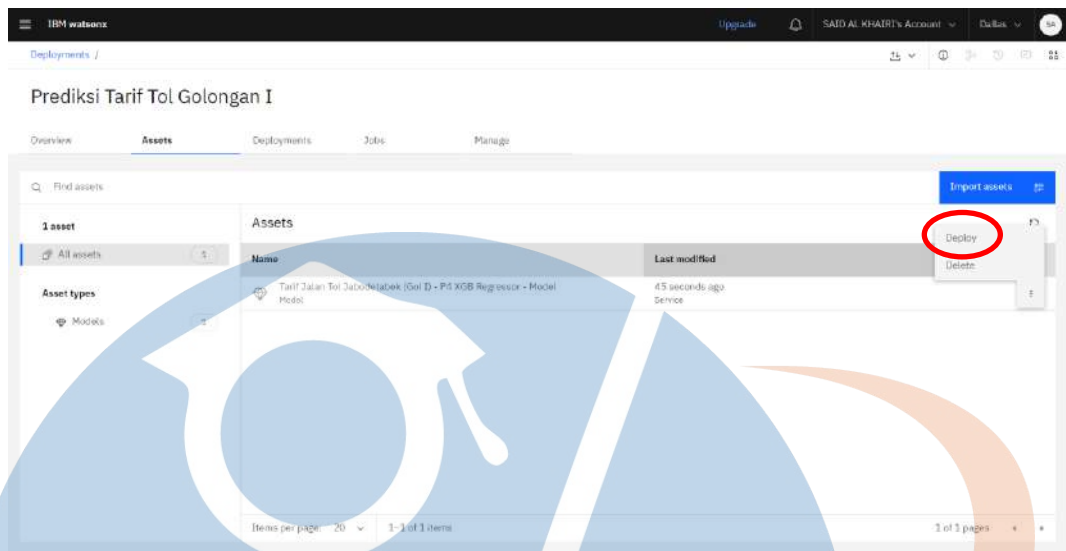
Gambar 30 Deploy Project

27. Pada bagian *machine learning service* → pilih *watson machine learning*. Lalu *save* untuk menyimpan.



Gambar 31 Manage Deployment

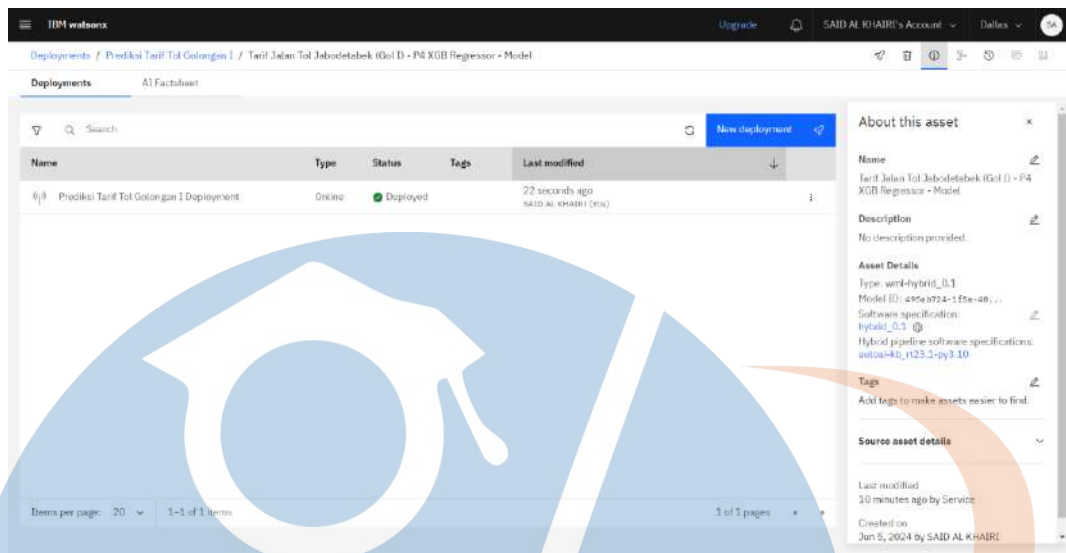
28. Kembali ke menu *assets*, lalu titik tiga pada model yang sebelumnya dibuat, kemudian pilih *deploy*.



Gambar 32 *Deploy Project*

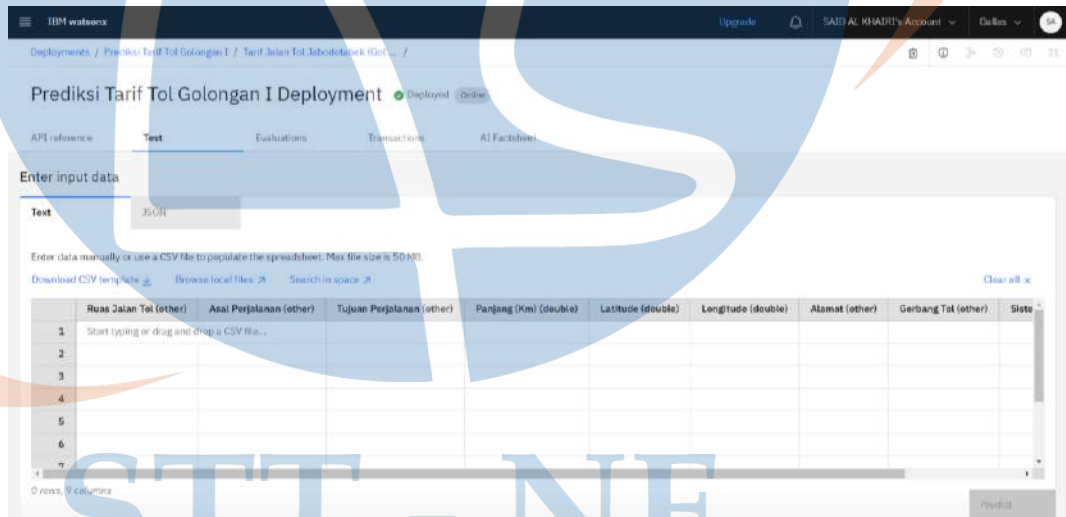
STT - NF

29. Setelah itu, pilih *Prediksi Tarif Tol Deployment* yang sebelumnya di *deploy*.



Gambar 33 Prediksi Tarif Tol

30. Pilih *test* untuk melakukan testing data yang ingin di tes. Jika sudah ada data pilih *predict*.



Gambar 34 Testing Project

4.3. 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, dimana 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.

4.3.1. Percobaan 1

Seperti yang dijelaskan sebelumnya percobaan pertama dilakukan pada kegiatan Studi Independen yang diselenggarakan oleh Infinite Learning. Saat percobaan pertama data yang digunakan adalah Jabodetabek-Banten, karena pada saat itu waktu yang diberikan hanya 2 minggu, jika memakai data tarif tol dari kaggle.com yang ditulis oleh Muhammad Yusuf Aristyanto terlalu banyak dan data tersebut masih berantakan.[21] Oleh karena itu, penulis menulis sendiri data tarif jalan tol Jabodetabek yang diterbitkan oleh Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat di website bpjt.pu.go.id. [19][20] Data tarif tol Jabodetabek terlalu sedikit dan batas minimum data yang bisa diolah adalah 100 baris, lalu penulis memutuskan untuk mengambil data tarif jalan tol Banten, jadilah data tarif jalan tol Jabodetabek-Banten. Lalu variabel yang diterbitkan oleh Badan Pengurus Jalan Tol hanya ruas jalan, asal perjalanan, tujuan perjalanan, dan tarif. Penulis menambahkan panjang jalan, karena panjang jalan menjadi salah satu faktor yang mempengaruhi tarif jalan tol. Jadi, penulis menggunakan asal perjalanan, tujuan perjalanan, panjang jalan, dan tarif (Gambar 35).

Dataset tersebut masih sangat sederhana, sehingga menghasilkan RMSE yang besar, yakni diatas 7000 bahkan ada yang mencapai 8000 dengan algoritma yang dipakai adalah *snap boosting machine regressor* (Gambar 36). Dan penulis baru mengetahui jika ada algoritma *snap boosting machine regressor*, karena sangat jarang sekali ada jurnal yang membahas algoritma ini, bahkan penulis sulit menemukan jurnal terkait algoritma ini. Cara kerja algoritma ini adalah dengan mengambil sampel subkelas secara acak pada

setiap iterasi boosting dan memilih hipotesis dasar yang meminimalkan fungsi kerugian untuk subkelas yang diambil sampelnya.

Dan dugaan penulis benar bahwa yang paling mempengaruhi hasil adalah panjang jalan yakni sebesar 100%, yang kedua asal perjalanan sebesar 24%, dan yang ketiga tujuan perjalanan sebesar 22% (Gambar 37). Penulis tidak terkejut karena data yang digunakan masihlah sederhana.

Dari hasil testing percobaan pertama, dimana hanya ada satu yang sesuai tarif prediksi dengan tarif asli, yaitu Pamulang - Serpong dimana kedua tabel tersebut sama - sama bernilai 11000. Dan ada yang menarik dari hasil tersebut dimana panjang jalan dan tarif sama tetapi hasil prediksi tersebut berbeda beda, ini menunjukkan asal perjalan dan tujuan perjalan itu berpengaruh seperti yang dijelaskan sebelumnya. Selain itu ada tarif prediksi yang nilainya dibawah tarif asli, tetapi lebih banyak tarif prediksi diatas tarif asli. Dan dari hasil testing tersebut selisih antara 500 – 7000 (Tabel 3). Jadi hasil testing pada percobaan pertama ini baik, mengingat akurasi yang ditunjukkan oleh metrik RMSE berada di 7775.567. Namun pada percobaan pertama masih bisa diperbaiki lagi karena fitur – fitur yang mempengaruhi jalan tol bukan hanya asal perjalanan, tujuan perjalanan, panjang jalan, dan golongan perjalanan saja.

4.3.2. Percobaan 2

Karena pada percobaan pertama penulis merasa akurasi yang dihasilkan masih besar, maka pada percobaan kedua penulis mencoba untuk menambahkan ruas jalan tol, *latitude*, *longitude*, alamat, gerbang tol, dan sistem pembayaran dengan harapan akurasi yang dihasilkan semakin kecil sehingga akan semakin akurat. Namun pada percobaan kedua menggunakan data tarif jalan tol Jabodetabek saja, karena gerbang tol yang ada di Jabodetabek ada lebih dari 200 gerbang tol. Jadi, data tarif jalan tol Jabodetabek saja sudah cukup.

Pada saat data diolah dan dipelajari nilai RMSE turun tetapi tidak signifikan. Nilai *top performance* RMSE pada percobaan kedua berada di angka 6607.068 dan algoritma yang digunakan berubah menjadi *snap random forest regressor* seperti yang terlihat pada gambar 4.34. Dimana algoritma ini

terdiri dari banyak pohon keputusan (*decision trees*) untuk mencapai hasil yang lebih akurat.[17] Algoritma ini bekerja dengan membangun beberapa pohon keputusan secara independen.[17] Selain algoritma *snap random forest regressor* yang menjadi *top performance* ada algoritma *xgb regressor* yang menjadi rank 3 dengan nilai RMSE diatas 7000 dan ada algoritma *ridge* yang menjadi rank 5 dengan nilai RMSE diatas 7000 pada percobaan kedua (Gambar 40).

Dan penulis terkejut karena variabel yang mempengaruhi hasil hanya asal perjalanan saja, dan itupun angka yang sangat kecil yakni sebesar 2.9% saja (Gambar 41). Penulis terheran - heran, karena pada percobaan kedua ini diharapkan lebih akurat dari percobaan pertama. Tetapi kenyataanya kurang lebih sama dengan percobaan pertama.

Pada percobaan kedua ini nilai akurasi yang ditunjukkan oleh metrik RMSE berada di 6607.068 yang mana lebih kecil sedikit dari percobaan pertama, namun pada hasil testing lebih buruk dari percobaan pertama. Meskipun begitu, penulis tetap testing model *machine learning* prediksi tarif jalan tol.

Dilihat dari hasil percobaan tersebut (Tabel 4) dimana dari testing percobaan kedua ini tidak ada tarif prediksi yang sesuai dengan tarif asli. Dan selisih hasil tarif prediksi dengan tarif asli pada percobaan kedua ini menjadi yang terbesar dari ketiga percobaan yang dilakukan, dimana selisih dari kedua tarif tersebut berkisar dari 1500 - 8000.

Jadi hasil testing pada percobaan pertama ini buruk, hasil testing ini terbilang buruk karena selisihnya sangat jauh dibandingkan dengan percobaan pertama mengingat akurasi yang ditunjukkan oleh metrik RMSE berada di 7775.567. Namun pada percobaan pertama masih bisa diperbaiki lagi karena fitur – fitur yang mempengaruhi jalan tol bukan hanya asal perjalanan, tujuan perjalanan, panjang jalan, dan golongan perjalanan saja.

4.3.3. Percobaan 3

Setelah diselidiki mengapa pada percobaan kedua tidak turun secara signifikan, akhirnya penulis menemukan masalahnya. Masalah itu terjadi karena kesalahan pada saat *convert* data dari *.xlsx* ke *.csv*. Pada data *.xlsx* latitude dan longitude terbaca *general*, tetapi saat *convert* ke *.csv* latitude dan

longitude terbaca *number* sehingga nilai latitude dan longitude menjadi jutaan hingga milyaran. Di percobaan ketiga ini penulis memperbaiki nilai – nilai yang salah diperbaiki (Gambar 4).

Pada saat data selesai diolah dan dipelajari (Gambar 19) ada yang menarik dari hasil tersebut, dimana model - model yang dihasilkan menggunakan *top performance* algoritma yang sama dengan percobaan sebelumnya. Berikut pembahasan lengkap ketiga algoritma tersebut.

4.3.3.1. *XGB Regressor*

Algoritma ini bekerja dengan membangun beberapa pohon keputusan secara berurutan, dengan setiap pohon mengoreksi kesalahan dari pohon sebelumnya, sehingga menghasilkan model prediksi yang lebih akurat.[14] algoritma *xgb regressor* sebagai *top performance* dengan nilai RMSE berada diangka 3390.691.

Pada *feature summary* (Gambar 21) menunjukkan faktor atau fitur yang berpengaruh, dimana *newfeature_15* yang bertransformasi data menggunakan *Principal Component Analysis (PCA)* pada seluruh set data dan menghasilkan komponen ke-8 mempengaruhi 50,24% dalam hasil prediksi. Setelah itu asal perjalanan mempengaruhi 20,14% dalam hasil prediksi, lalu *newfeature_10* yang bertransformasi data menggunakan *Principal Component Analysis (PCA)* pada seluruh set data dan menghasilkan komponen ke-1 mempengaruhi 12,79% dalam hasil prediksi. Kemudian panjang jalan mempengaruhi 3,43% dalam hasil prediksi, dan sisanya hanya dibawah 3%.

Dari penjelasan sebelumnya fitur original yang mempengaruhi hasil prediksi pada algoritma ini ada asal perjalanan sebesar 20,14%, panjang jalan sebesar 3,43%, dan tujuan perjalanan sebesar 0,89% (Gambar 21).

Dan dari hasil testing pada algoritma ini dimana ada 7 yang sesuai tarif prediksi dengan tarif asli, yaitu Serpong – Simpang Susun CBD, Junction Benda – Tanah Tinggi, Sentul Selatan – Selatan Semplak, Depok – Antasari, Pondok Pinang – Taman mini, Casablanca – Pondok Kelapa, Cibitung – Gabus. Selain itu ada tarif prediksi yang nilainya dibawah tarif asli, tetapi kebanyakan tarif prediksi diatas tarif asli. Dan juga ada beberapa perbedaan

antara tarif prediksi dengan tarif asli di atas 5000 seperti Jakarta – Bogor, Tomang IC – Tangerang, Cawang – Tanjung Priok, Cawang – Tanjung Priok, Kelapa Gading – Pulo Gebang, Junction Cimanggis – Simpang Susun Cikeas. Namun jika dilihat dari fitur pendukungnya hal tersebut wajar misalnya Junction Cimanggis – Simpang Susun Cikeas tarif asli 13500 tarif prediksi 6000 tetapi ada panjang jalan sepanjang 7 km. Jadi hasil testing pada percobaan pertama ini baik, namun pada algoritma ini selisih tarif asli dengan tarif prediksi yang berada di atas 5000 ada banyak dan menjadi yang terbanyak dari ketiga algoritma yang dilakukan testing.

4.3.3.2. *Snap Boosting Machine Regressor*

Algoritma ini bekerja dengan mengambil sampel subkelas secara acak pada setiap iterasi boosting dan memilih hipotesis dasar yang meminimalkan fungsi kerugian untuk subkelas yang diambil sampelnya.[16] Algoritma *snap boosting machine regressor* yang berada di urutan kedua setelah *top performance* dan juga urutan kedua dalam algoritma, dengan nilai RMSE diangka 4027.684.

Pada *feature summary* (Gambar 43) menunjukkan faktor atau fitur yang berpengaruh, dimana *newfeature_8* yang bertransformasi data menggunakan *Principal Component Analysis (PCA)* pada seluruh set data dan menghasilkan komponen ke-6 yang dikalikan oleh data panjang(km) mempengaruhi sebesar 54,45% dalam hasil prediksi. Setelah itu asal perjalanan mempengaruhi sebesar 14,23% dalam hasil prediksi, lalu panjang(km) mempengaruhi sebesar 8,63%. Selanjutnya adalah fitur *newfeature_2* yang bertransformasi data menggunakan *Principal Component Analysis (PCA)* pada seluruh set data dan menghasilkan komponen ke-4 mempengaruhi sebesar 7,72%. Lalu ada *newfeature_23* yang bertransformasi data menggunakan *principal component analysis (PCA)* pada seluruh set data dan menghasilkan komponen ke-4 yang dilalikan oleh data menggunakan *principal component analysis (PCA)* pada seluruh set data dan menghasilkan komponen ke-3 mempengaruhi sebesar 5,66%. Kemudian ruas jalan tol mempengaruhi sebesar 4,36%, lalu ada *newfeature_3* yang bertransformasi data

menggunakan *Principal Component Analysis (PCA)* pada seluruh set data dan menghasilkan komponen ke-6 mempengaruhi sebesar 3,44%.

Dari penjelasan sebelumnya faktor atau fitur original yang berpengaruh pada algoritma *snap boosting machine regressor* ini adalah asal perjalanan sebesar 14,23%, lalu panjang jalan sebesar 8,63%, dan terakhir ruas jalan tol mempengaruhi sebesar 4,36%.

Pada dasarnya hasil testing (Tabel 6) pada algoritma ini tidak jauh berbeda dengan algoritma *xgb regressor* dimana ada 7 yang sesuai tarif prediksi dengan tarif asli, yaitu Akses Tanjung Priok Seksi E-2A - Akses Tanjung Priok Seksi E-2, Serpong - Simpang Susun CBD, Pondok Aren - Serpong, Junction Serpong - Pamulang, Junction Benda - Tanah Tinggi, Depok - Antasari, dan Pondok Pinang - Taman mini. Namun hasil testing pada algoritma ini lebih baik dari algoritma *xgb regressor (top performance)*, karena selisih tarif asli dengan tarif prediksi relatif kecil dan yang diatas 5000 hanya ada satu.

4.3.3.3. *Random Forest Regressor*

Algoritma ini terdiri dari banyak pohon keputusan (*decision trees*) untuk mencapai hasil yang lebih akurat.[17] Algoritma ini bekerja dengan membangun beberapa pohon keputusan secara independen.[17] Algoritma *random forest regressor* yang berada di urutan kelima dan urutan ketiga dalam algoritma, dengan nilai RMSE diangka 4554.243.

Faktor atau fitur original yang berpengaruh pada algoritma *snap boosting machine regressor* ini (Gambar 44) adalah panjang jalan sebesar 31,76%, kemudian asal perjalanan sebesar 10,11%, lalu latitude sebesar 7,51%, dan terakhir longitude mempengaruhi sebesar 7,26%.

Dan dari hasil testing (Tabel 7) pada algoritma ini dimana ada 3 yang sesuai tarif prediksi dengan tarif asli, Pluit - Cengkareng, Cawang - Tomang, dan Sentul Selatan - Selatan Semplak. Hasil testing pada algoritma ini terbilang baik, karena selisih tarif asli dengan tarif prediksi 500 - 9000. Ini merupakan selisih terkecil dari ketiga algoritma yang di testing.

Jika dilihat perbedaan dari ketiga algoritma tersebut pada tabel berikut.

Tabel 2 Perbandingan Ketiga Algoritma

Algoritma	<i>Extreme Gradient Boosting Regressor</i>	<i>Snap Boosting Machine Regressor</i>	<i>Random Forest Regressor</i>
Data	Jabodetabek	Jabodetabek	Jabodetabek
Akurasi	<i>Root Mean Square Error</i> = 3390.691	<i>Root Mean Square Error</i> = 4027.684	<i>Root Mean Square Error</i> = 4554.243
Fitur Berpengaruh	Asal Perjalan 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 tarif asli dengan tarif 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 perjalan sebesar 20,14%, panjang jalan sebesar 3,43%, dan tujuan perjalanan sebesar 0,89% (Gambar 21). 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* dimana 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.



STT - NF

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan penelitian yang telah dipaparkan pada bab sebelumnya, peneliti dapat mengambil kesimpulan diantaranya:

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, dimana asal perjalanan di angka 20,14%, panjang jalan di angka 3,43%, dan tujuan perjalanan di angka 0,89%.

5.2 Saran

Berdasarkan penelitian yang dilakukan, penulis memberikan saran yang dapat diimplementasikan pada penelitian selanjutnya, yaitu sebagai berikut:

- 1 Pada penelitian selanjutnya disarankan untuk melakukan pembahasan cara mengimplementasikan model *machine learning* tersebut menjadi berbasis website.

STT - NF

DAFTAR PUSTAKA

- [1] Indonesia, "Undang - Undang Republik Indonesia Nomor 38 Tahun 2004 Tentang Jalan," 2004.
- [2] Indonesia, "Peraturan Menteri Pekerjaan Umum dan Perumahan Rakyat (PUPR) No. 16/PRT/M/2017," 2017.
- [3] Indonesia, "Peraturan Pemerintah (PP) Nomor 15 Tahun 2005 Tentang Jalan Tol," 2005.
- [4] Sistem Informasi Geografis Badan Pengatur Jalan Tol, "Peta Jalan Tol Indonesia," bpjt.pu.go.id . Diakses: 14 Maret 2024. [Daring]. Tersedia pada:
<https://sigi.pu.go.id/portalpupr/apps/dashboards/ad691982b770462d8e236f8ca7e450f4>
- [5] Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat, "Monitoring," bpjt.pu.go.id . Diakses: 3 Februari 2024. [Daring]. Tersedia pada: <https://bpjt.pu.go.id/konten/monitoring/jabodetabek>
- [6] Rosalita L, Purba A, dan 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.
- [7] Indonesia, "Peraturan Pemerintah (PP) Nomor 30 Tahun 2017 Tentang Perubahan Ketiga atas Peraturan Pemerintah Nomor 15 Tahun 2005 tentang Jalan Tol," 2017.
- [8] Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat, "Golongan Kendaraan," bpjt.pu.go.id . Diakses: 20 Juni 2024. [Daring]. Tersedia pada: <https://bpjt.pu.go.id/golongan-kendaraan>
- [9] J. Syahputra, R. Dias Ramadhani, dan A. Burhanudin, "Prediksi Harga Saham Bank Bri Menggunakan Algoritma Linear Regresion Sebagai Strategi Jual Beli Saham," *Data Institut Teknologi Telkom Purwokerto*, vol. 2, no. 1, hlm. 1–10, 2022, [Daring]. Tersedia pada: <http://journal.itelkom-pwt.ac.id/index.php/dinda>
- [10] Mikhael, F. Andreas, dan U. Enri, "Perbandingan Algoritma Linier Regression, Neural Network, Deep Learning, Dan K-Nearest Neighbor (K-NN) Untuk Prediksi Bitcoin," *JSI : Jurnal Sistem Informasi (E-Journal)*, vol. 14, no. 1, 2022, [Daring]. Tersedia pada: <http://ejournal.unsri.ac.id/index.php/jsi/index>
- [11] R. Agung Prasetyo dan Helma, "Analisis Regresi Linear Berganda Untuk Melihat Faktor Yang Berpengaruh Terhadap Kemiskinan di Provinsi Sumatera Barat," *Journal Of Mathematics UNP*, vol. 7, no. 2, Jul 2022.
- [12] Harsiti, Z. Muttaqin, dan E. Srihartini, "Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet," *Sistem Informasi*, vol. 9, no. 1, hlm. 12–16.
- [13] P. Kevin, "Regresi: Pengertian, Macam, Rumus, dan Contoh Soal," rumuspintar.com. Diakses: 21 April 2024. [Daring]. Tersedia pada: <https://rumuspintar.com/regresi/#:~:text=Regresi%20adalah%20suatu%20>

- metode%20statistik%20dengan%20merumuskan%20persamaan,prediksi
%20dengan%20data-
data%20yang%20diolah%20merupakan%20data%20kuantitatif
- [14] B. Jange, "Prediksi Harga Saham Bank BCA Menggunakan XGBoost," *ARBITRASE: Journal of Economics and Accounting*, vol. 3, no. 2, hlm. 231–237, Nov 2022, doi: 10.47065/arbitrase.v3i2.495.
- [15] J. Avanija, G. Sunitha, K. Reddy Madhavi, P. Kora, dan R. Vittal Sai Hitesh, "Prediction of House Price Using XGBoost Regression Algorithm," 2021.
- [16] T. Parnell *dkk.*, "SnapBoost: A Heterogeneous Boosting Machine".
- [17] S. Zahedian, A. Nohekhan, dan 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.
- [18] N. N. Hegde, S. Divakar Goudar, Varsha S, dan Nethravathi B, "Overview On IBM Cloud," *International Research Journal of Modernization in Engineering Technology and Science*, vol. 6, no. 1, Jan 2024, doi: 10.56726/irjmets48481.
- [19] Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat, "Cek Tarif Tol," bpjt.pu.go.id. Diakses: 3 Februari 2024. [Daring]. Tersedia pada: <https://bpjt.pu.go.id/cek-tarif-tol>
- [20] Badan Pengatur Jalan Tol Kementerian Pekerjaan Umum dan Perumahan Rakyat, "Tarif Tol," bpjt.pu.go.id. Diakses: 3 Februari 2024. [Daring]. Tersedia pada: <https://bpjt.pu.go.id/tabel-tarif-tol>
- [21] M. Yusuf Aristyanto, "Indonesian Toll Road Rates in June 2020," [kaggle.com](https://www.kaggle.com).
- [22] S. Arikunto, *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: Rineka Cipta, 2019.
- [23] Sugiyono, *Metode penelitian kuantitatif, kualitatif, dan R&D*. Bandung: Alfabeta, 2020.

STT - NF

LAMPIRAN

Lampiran 1

Asal Perjalanan	Tujuan Perjalanan	Panjang Jalan (± km)	Tarif
2 Seksi E-1, E-2, E-2A, NS dan NS (Rorotan-Kebon Bawang)	Seksi E-1, E-2, E-2A, NS dan NS (Rorotan-Kebon Bawang)	11	17000
3 Jakarta - Bogor - Ciawi	Jakarta - Bogor - Ciawi	59	7500
4 Tomang IC	Tangerang	26	8000
5 Tomang IC	Cikupa	26	8000
6 Tangerang	Tomang IC	26	8000
7 Tangerang	Cikupa	26	8000
8 Cikupa	Tomang IC	26	8000
9 Cikupa	Tangerang	26	8000
10 Serpong	Simpang Susun CBD	5	5500
11 Serpong	Panulang	7	11000
12 Panulang	Serpong	7	11000
13 Prof. Dr. H. Sedyatmo	Prof. Dr. H. Sedyatmo	14	8500
14 Jakarta IC	Pondok Gede Barat/Pondok Gede Timur	8	4000
15 Jakarta IC	Cikunir	10	7000
16 Jakarta IC	Bekasi Barat	13	7000
17 Jakarta IC	Bekasi Timur	16	7000
18 Jakarta IC	Tambun	21	7000
19 Jakarta IC	Cibitung	24	7000
20 Jakarta IC	Cikarang Barat	31	7000
21 Jakarta IC	Cibatu	34	12000
22 Jakarta IC	Cikarang Barat	37	12000
23 Pondok Aren	Serpong	7	7000

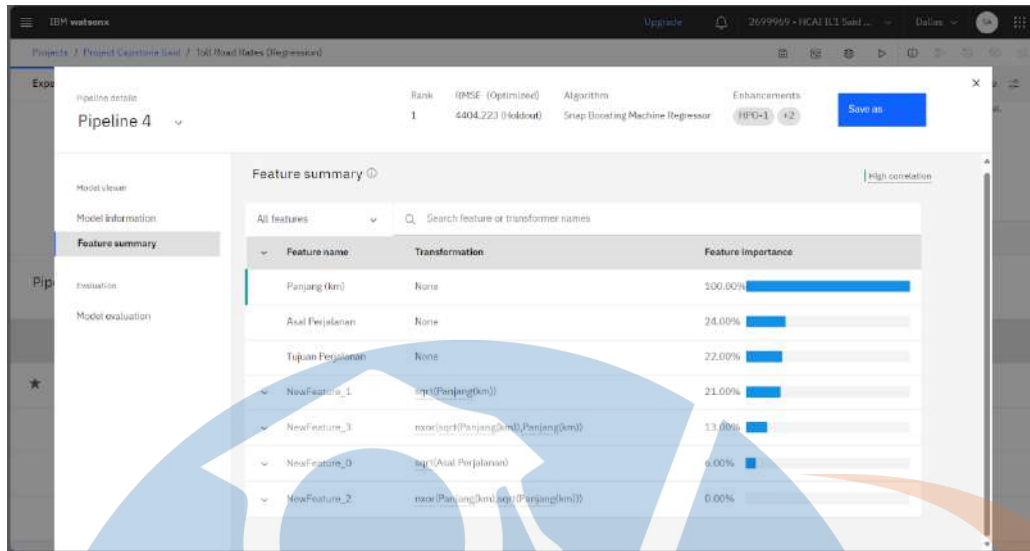
Gambar 35 Data Percobaan 1



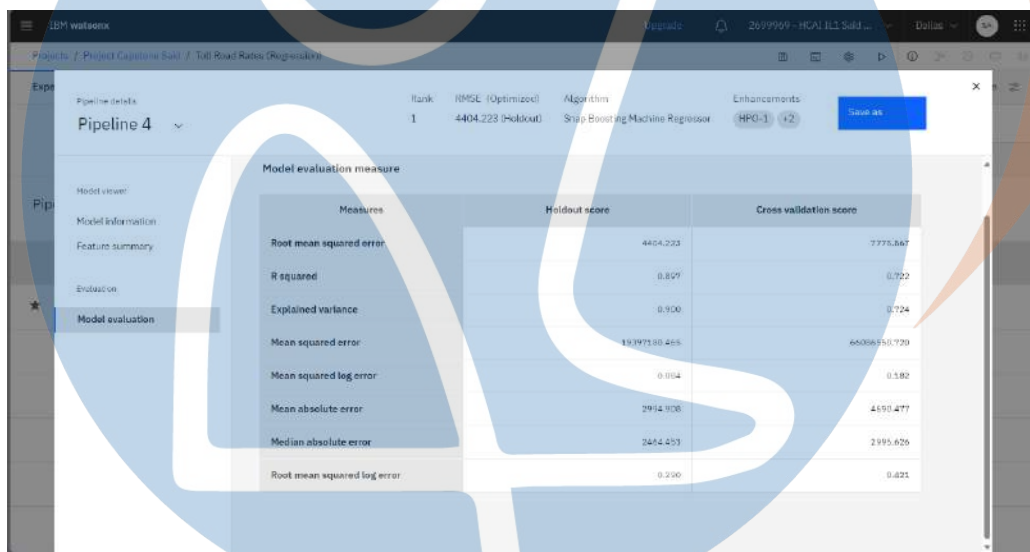
Pipeline leaderboard

Rank	Name	Algorithm	RMSE (Optimized) cross validation	Enhancements	Build time
★ 1	Pipeline 4	Snap Boosting Machine Regressor	7775.567	HPO-3 FE HPO-3	00:00:14
2	Pipeline 3	Snap Boosting Machine Regressor	7862.250	HPO-1 FE	00:00:12
3	Pipeline 2	Snap Boosting Machine Regressor	8093.775	HPO-2	00:00:01
4	Pipeline 1	Snap Boosting Machine Regressor	8530.012	None	00:00:01

Gambar 36 Hasil Percobaan 1



Gambar 37 Feature Summary Percobaan 1



Gambar 38 Model Evaluation

Tabel 3 Data Hasil Prediksi Percobaan 1

Asal Perjalanan	Tujuan Perjalanan	Panjang (km)	Tarif	Prediction
Jakarta - Bogor - Ciawi	Jakarta - Bogor - Ciawi	59	7500	11500
Pondok Aren	Serpong	7	7000	8500
Pondok Aren	Ulujami	9	16000	12000

Bintaro Viaduct	Pondok ranji	3	3000	3500
Tomang IC	Cikupa	26	8000	9000
Tangerang	Tomang IC	26	8000	7500
Tangerang	Cikupa	26	8000	11500
Cikupa	Tomang IC	26	8000	9500
Cikupa	Tangerang	26	8000	15000
Serpong	Simpang Susun CBD	5	5500	8500
Serpong	Pamulang	7	11000	11500
Pamulang	Serpong	7	11000	11000

STT - NF

Lampiran 2

Jabodetabek Toll Road Rates (Gol I)

File Edit Tampilan Sisipkan Format Data Alat Ekstensi Bantuan

100% 123 Roboto

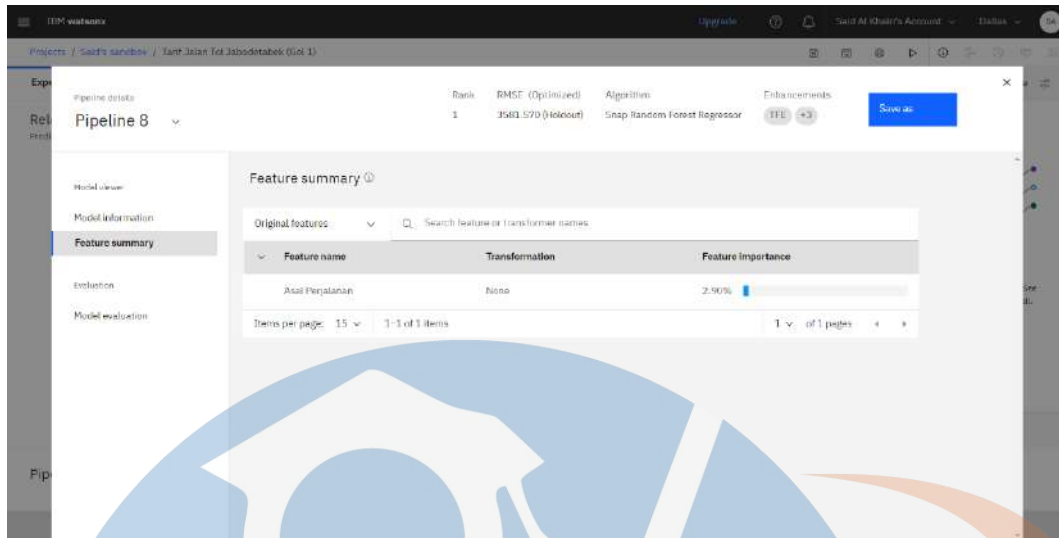
A1	A	B	C	D	E	F
1	Ruas Jalan Tol	Asal Perjalanan	Tujuan Perjalanan	Panjang (Krn)	Latitude	Longitude
2	Akses Tanjung Priok	Akses Tanjung Priok Seksi E-1	Akses Tanjung Priok Seksi E-2	0,40	-6,1195274	106,8932664
3	Akses Tanjung Priok	Akses Tanjung Priok Seksi E-2	Akses Tanjung Priok Seksi NS	2,74	-6,1084388	106,9025016
4	Akses Tanjung Priok	Akses Tanjung Priok Seksi E-2A	Akses Tanjung Priok Seksi E-2	1,92	-6,1040771	106,8992844
5	Akses Tanjung Priok	Akses Tanjung Priok Seksi NS	Akses Tanjung Priok Seksi E-2A	2,24	-6,1083637	106,9124905
6	Akses Tanjung Priok	Akses Tanjung Priok Seksi NS (Rorotan-Kebon Bawang)	Akses Tanjung Priok Seksi E-2A	1,10	-6,1179397	106,9255067
7	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,2871125	106,8762589
8	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,2875398	106,8776752
9	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,2960156	106,8783167
10	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3024932	106,8830419
11	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,2906532	106,8756233
12	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3101479	106,8847408
13	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3560822	106,8797922
14	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3652993	106,8942298
15	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3549369	106,8428572
16	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3827938	106,8957389
17	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3866476	106,8954882
18	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3852023	106,8955138
19	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,3876732	106,896188
20	Jakarta - Bogor - Ciawi	Jakarta	Bogor	19,00	-6,4612768	106,8893438
21	Jakarta - Bogor - Ciawi	Bogor	ciawi	23,00	-6,4572121	106,8840701
22	Jakarta - Bogor - Ciawi	Bogor	ciawi	23,00	-6,4824656	106,8744362

Das Akhir Revisi Data Project Capstone Data Cadangan Data Target Cadangan Data Mentah ezGeocode #5

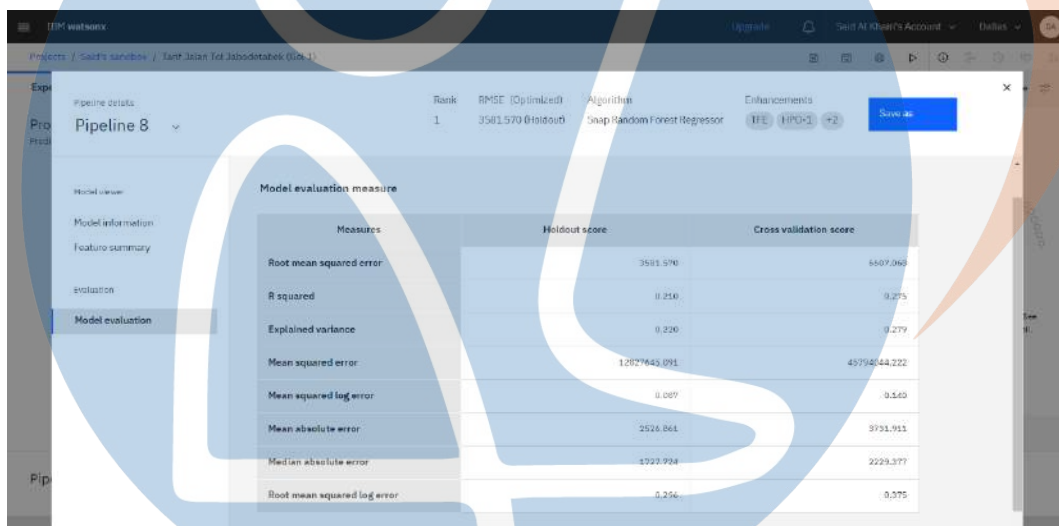
Gambar 39 Data Percobaan 2



Gambar 40 Hasil Percobaan 2



Gambar 41 Feature Summary Percobaan 2



Gambar 42 Model Evaluation

Tabel 4 Data Hasil Prediksi Percobaan 2

Asal Perjalanan	Tujuan Perjalanan	Panjang (Km)	Gerbang Tol	Tarif	Prediksi
Akses Tanjung Priok Seksi E-2A	Akses Tanjung Priok Seksi E-2	1,92	Gerbang Tol Koja Direct	17000	14000
Jakarta	Bogor	19,00	Gerbang Tol Gunung Putri	7500	13000
Simpang Susun Ciawi	Ciawi	1,00	Gerbang Tol Caringin	7500	13500
Tomang IC	Tangerang	18,00	Gerbang Tol Tangerang 2	8000	14500

Serpong	Simpang Susun CBD	5,00	Gerbang Tol BSD Timur 1	5500	12000
Pluit	Cengkareng	12,00	Gerbang Tol Kapuk	8500	14000
Jakarta IC	Tambun	21,00	Gerbang Tol Tambun	9500	14500
Junction Kunciran	Simpang Susun Parigi	7,00	Gerbang Tol Parigi	12500	14000
Pondok Aren	Serpong	7,00	Gerbang Pondok Aren Barat 2	7000	14000
Junction Serpong	Pamulang	7,00	Gerbang Tol Pamulang	12000	15000
Bintaro	Pondok Ranji	3,00	Gerbang Tol Bintaro 2	7000	14500
Junction Benda	Tanah Tinggi	7,00	Gerbang Tol Tanah Tinggi 1	15500	14000
Cawang	Tomang	14	Gerbang Tol Senayan	10500	14000
Cawang	Tanjung Priok	12,9	Gerbang Tol Pulomas	10500	15000
Ancol Timur	Jembatan Tiga	8,00	Gerbang Tol Kemayoran	10500	13000
Kelapa Gading	Pulo Gebang	9,00	Gerbang Tol Cakung (JTD)	22000	14000
Sentul Selatan	Selatan Semplak	14,00	Gerbang Tol Sentul Barat	15000	13000
Depok	Antasari	21,54	Gerbang Tol Brigif 1	13500	14000
Junction Cimanggis	Simpang Susun Cikeas	7	Gerbang Tol Nagrak (Kota Wisata)	13500	15000
Cakung	Cilingcing	4,08	Gerbang Tol Cakung 2	17000	18500
Cikunir	Cakung	9,7	Gerbang Tol Bintara	17000	15000
Taman Mini	Cikunir	12,4	Gerbang Tol Jati Asih 1	17000	13500

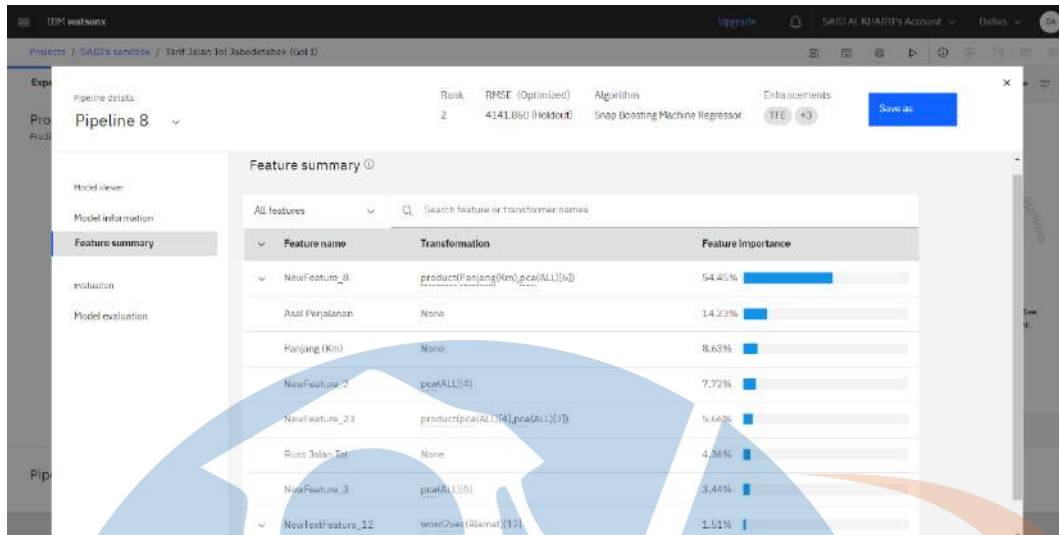
Pondok Pinang	Taman mini	12,9	Gerbang Tol Ampera 2	17000	16000
Kembangan	Pondok Pinang	11,21	Gerbang Tol Joglo 3	17000	14500
Penjaringan	Kembangan	8,65	Gerbang Tol Rawa Buaya Selatan	17000	14000
Casablanca	Pondok Kelapa	6,4	Gerbang Tol Pondok Kelapa 1	16000	14500
Cibitung	Gabus	12,00	Gerbang Tol Gabus	25500	19000

Lampiran 3

Tabel 5 Data Hasil Prediksi Percobaan 3 (*XGB Regressor*)

Asal Perjalanan	Tujuan Perjalanan	Panjang (Km)	Gerbang Tol	Tarif	Prediksi
Akses Tanjung Priok Seksi E-2A	Akses Tanjung Priok Seksi E-2	1,92	Gerbang Tol Koja Direct	17000	13500
Jakarta	Bogor	19,00	Gerbang Tol Gunung Putri	7500	19000
Simpang Susun Ciawi	Ciawi	1,00	Gerbang Tol Caringin	7500	4000
Tomang IC	Tangerang	18,00	Gerbang Tol Tangerang 2	8000	23500
Serpong	Simpang Susun CBD	5,00	Gerbang Tol BSD Timur 1	5500	5500
Pluit	Cengkareng	12,00	Gerbang Tol Kapuk	8500	7000
Jakarta IC	Tambun	21,00	Gerbang Tol Tambun	9500	13000
Junction Kunciran	Simpang Susun Parigi	7,00	Gerbang Tol Parigi	12500	12000
Pondok Aren	Serpong	7,00	Gerbang Pondok Aren Barat 2	7000	8000
Junction Serpong	Pamulang	7,00	Gerbang Tol Pamulang	12000	9500

Bintaro	Pondok Ranji	3,00	Gerbang Tol Bintaro 2	7000	10500
Junction Benda	Tanah Tinggi	7,00	Gerbang Tol Tanah Tinggi 1	15500	15500
Cawang	Tomang	14	Gerbang Tol Senayan	10500	8000
Cawang	Tanjung Priok	12,9	Gerbang Tol Pulomas	10500	24000
Ancol Timur	Jembatan Tiga	8,00	Gerbang Tol Kemayoran	10500	17000
Kelapa Gading	Pulo Gebang	9,00	Gerbang Tol Cakung (JTD)	22000	16500
Sentul Selatan	Selatan Semplak	14,00	Gerbang Tol Sentul Barat	15000	15000
Depok	Antasari	21,54	Gerbang Tol Brigif 1	13500	13500
Junction Cimanggis	Simpang Susun Cikeas	7	Gerbang Tol Nagrak (Kota Wisata)	13500	6000
Cakung	Cilingcing	4,08	Gerbang Tol Cakung 2	17000	15500
Cikunir	Cakung	9,7	Gerbang Tol Bintara	17000	16500
Taman Mini	Cikunir	12,4	Gerbang Tol Jati Asih 1	17000	20000
Pondok Pinang	Taman mini	12,9	Gerbang Tol Ampera 2	17000	17000
Kembangan	Pondok Pinang	11,21	Gerbang Tol Joglo 3	17000	16500
Penjaringan	Kembangan	8,65	Gerbang Tol Rawa Buaya Selatan	17000	18000
Casablanca	Pondok Kelapa	6,4	Gerbang Tol Pondok Kelapa 1	16000	16000
Cibitung	Gabus	12,00	Gerbang Tol Gabus	25500	25500

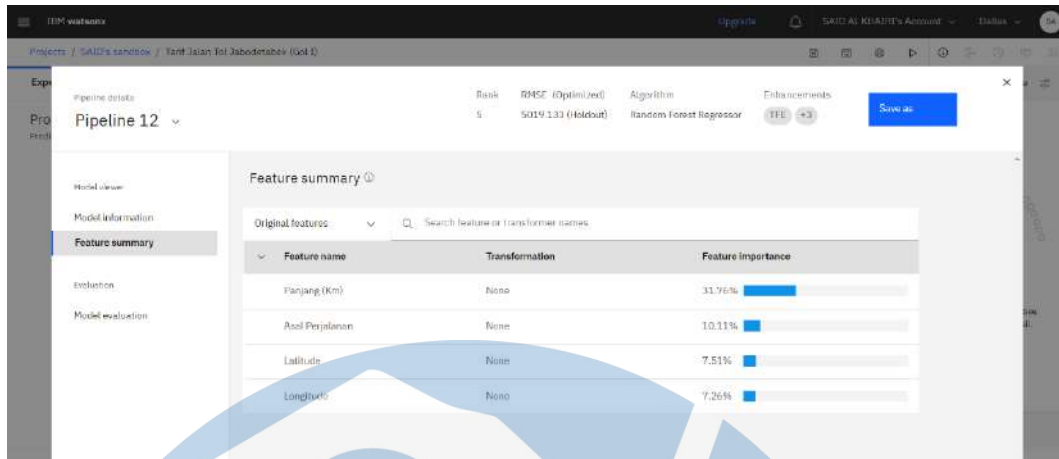


Gambar 43 Feature Summary Percobaan 3 (Snap Boosting Machine Regressor)

Tabel 6 Data Hasil Prediksi Percobaan 3 (Snap Boosting Machine Regressor)

Asal Perjalanan	Tujuan Perjalanan	Panjang (Km)	Gerbang Tol	Tarif	Prediksi
Akses Tanjung Priok Seksi E-2A	Akses Tanjung Priok Seksi E-2	1,92	Gerbang Tol Koja Direct	17000	17000
Jakarta	Bogor	19,00	Gerbang Tol Gunung Putri	7500	10500
Simpang Susun Ciawi	Ciawi	1,00	Gerbang Tol Caringin	7500	6500
Tomang IC	Tangerang	18,00	Gerbang Tol Tangerang 2	8000	11000
Serpong	Simpang Susun CBD	5,00	Gerbang Tol BSD Timur 1	5500	5500
Pluit	Cengkareng	12,00	Gerbang Tol Kapuk	8500	24000
Jakarta IC	Tambun	21,00	Gerbang Tol Tambun	9500	10000
Junction Kunciran	Simpang Susun Parigi	7,00	Gerbang Tol Parigi	12500	13500
Pondok Aren	Serpong	7,00	Gerbang Pondok Aren Barat 2	7000	7000
Junction Serpong	Pamulang	7,00	Gerbang Tol Pamulang	12000	12000

Bintaro	Pondok Ranji	3,00	Gerbang Tol Bintaro 2	7000	11500
Junction Benda	Tanah Tinggi	7,00	Gerbang Tol Tanah Tinggi 1	15500	15500
Cawang	Tomang	14	Gerbang Tol Senayan	10500	9000
Cawang	Tanjung Priok	12,9	Gerbang Tol Pulomas	10500	10000
Ancol Timur	Jembatan Tiga	8,00	Gerbang Tol Kemayoran	10500	8000
Kelapa Gading	Pulo Gebang	9,00	Gerbang Tol Cakung (JTD)	22000	20000
Sentul Selatan	Selatan Semplak	14,00	Gerbang Tol Sentul Barat	15000	14500
Depok	Antasari	21,54	Gerbang Tol Brigif 1	13500	13500
Junction Cimanggis	Simpang Susun Cikeas	7	Gerbang Tol Nagrak (Kota Wisata)	13500	13000
Cakung	Cilingcing	4,08	Gerbang Tol Cakung 2	17000	17500
Cikunir	Cakung	9,7	Gerbang Tol Bintara	17000	17500
Taman Mini	Cikunir	12,4	Gerbang Tol Jati Asih 1	17000	15000
Pondok Pinang	Taman mini	12,9	Gerbang Tol Ampera 2	17000	17000
Kembangan	Pondok Pinang	11,21	Gerbang Tol Joglo 3	17000	18000
Penjaringan	Kembangan	8,65	Gerbang Tol Rawa Buaya Selatan	17000	14500
Casablanca	Pondok Kelapa	6,4	Gerbang Tol Pondok Kelapa 1	16000	13000
Cibitung	Gabus	12,00	Gerbang Tol Gabus	25500	27000



Gambar 44 Feature Summary Percobaan 3 (Random Forest Regressor)

Tabel 7 Data Hasil Prediksi Percobaan 3 (Random Forest Regressor)

Asal Perjalanan	Tujuan Perjalanan	Panjang (Km)	Gerbang Tol	Tarif	Prediksi
Akses Tanjung Priok Seksi E-2A	Akses Tanjung Priok Seksi E-2	1,92	Gerbang Tol Koja Direct	17000	13500
Jakarta	Bogor	19,00	Gerbang Tol Gunung Putri	7500	15000
Simpang Susun Ciawi	Ciawi	1,00	Gerbang Tol Caringin	7500	8500
Tomang IC	Tangerang	18,00	Gerbang Tol Tangerang 2	8000	9000
Serpong	Simpang Susun CBD	5,00	Gerbang Tol BSD Timur 1	5500	7500
Pluit	Cengkareng	12,00	Gerbang Tol Kapuk	8500	8500
Jakarta IC	Tambun	21,00	Gerbang Tol Tambun	9500	14500
Junction Kunciran	Simpang Susun Parigi	7,00	Gerbang Tol Parigi	12500	9000
Pondok Aren	Serpong	7,00	Gerbang Pondok Aren Barat 2	7000	7500
Junction Serpong	Pamulang	7,00	Gerbang Tol Pamulang	12000	9000

Bintaro	Pondok Ranji	3,00	Gerbang Tol Bintaro 2	7000	12000
Junction Benda	Tanah Tinggi	7,00	Gerbang Tol Tanah Tinggi 1	15500	15000
Cawang	Tomang	14	Gerbang Tol Senayan	10500	10500
Cawang	Tanjung Priok	12,9	Gerbang Tol Pulomas	10500	10000
Ancol Timur	Jembatan Tiga	8,00	Gerbang Tol Kemayoran	10500	9000
Kelapa Gading	Pulo Gebang	9,00	Gerbang Tol Cakung (JTD)	22000	16000
Sentul Selatan	Selatan Semplak	14,00	Gerbang Tol Sentul Barat	15000	15000
Depok	Antasari	21,54	Gerbang Tol Brigif 1	13500	15000
Junction Cimanggis	Simpang Susun Cikeas	7	Gerbang Tol Nagrak (Kota Wisata)	13500	9000
Cakung	Cilingcing	4,08	Gerbang Tol Cakung 2	17000	15000
Cikunir	Cakung	9,7	Gerbang Tol Bintara	17000	16000
Taman Mini	Cikunir	12,4	Gerbang Tol Jati Asih 1	17000	15500
Pondok Pinang	Taman mini	12,9	Gerbang Tol Ampera 2	17000	16000
Kembangan	Pondok Pinang	11,21	Gerbang Tol Joglo 3	17000	16000
Penjaringan	Kembangan	8,65	Gerbang Tol Rawa Buaya Selatan	17000	16000
Casablanca	Pondok Kelapa	6,4	Gerbang Tol Pondok Kelapa 1	16000	16500
Cibitung	Gabus	12,00	Gerbang Tol Gabus	25500	16500