

Implementasi *Machine Learning* pada Sistem Prediksi Kejadian dan Lokasi Patah Rel Kereta Api di Indonesia

Nafisah Nurul Hakim

Sekolah Teknik Elektro dan Informatika

Institut Teknologi Bandung

Bandung, Indonesia

nafisah.nurul15@gmail.com

Abstrak—Besarnya minat masyarakat Indonesia terhadap penggunaan kereta api menyebabkan adanya urgensi terhadap penjaminan keselamatan penggunaan kereta api yang baik dari penyelenggaranya. Meskipun begitu, KNKT menemukan bahwa sebagian besar kecelakaan kereta api disebabkan oleh faktor prasarana, salah satunya adalah rel patah. Peninjauan lebih lanjut pada kejadian terkait memberikan hasil bahwa kejadian rel patah ini disebabkan oleh beberapa hal, di antaranya terdapat pada aspek manajemen dan organisasi. Penelitian ini dilakukan untuk memberikan pandangan baru terhadap metode penyelesaian masalah pada aspek manajemen dan organisasi dengan menerapkan *machine learning* untuk melakukan penilaian kondisi rel. Metode penilaian yang dimaksud dalam penelitian ini adalah penggunaan model *machine learning* untuk melakukan prediksi kemungkinan terjadinya rel patah di suatu titik dengan kondisi-kondisi yang dimasukkan sebagai *input* pada model. Pengembangan model dilakukan menggunakan metodologi CRISP-DM dan beberapa teknik pemodelan yang saling dibandingkan hasilnya sehingga menghasilkan model paling tepat. Evaluasi pada hasil setiap model yang dilakukan di akhir penelitian memberikan kesimpulan bahwa *random forest* adalah teknik yang paling tepat untuk digunakan dalam melakukan pembuatan model analisis prediksi kejadian dan lokasi rel patah berdasarkan data yang digunakan.

Keywords—kereta api, rel patah, *machine learning*, CRISP-DM

I. PENDAHULUAN

Kereta api merupakan salah satu moda transportasi massal yang masih cukup diminati masyarakat hingga saat ini [1]. Salah satu penyebabnya adalah penggunaan kereta api yang membuat perjalanan menghabiskan waktu secara lebih efektif dibandingkan dengan penggunaan mode transportasi lain. Selain itu, kereta api termasuk mode transportasi publik yang hingga saat ini biayanya cukup terjangkau oleh semua level masyarakat dengan berbagai jumlah penghasilan [2].

Besarnya minat masyarakat terhadap kereta api dapat dibuktikan dengan meningkatnya jumlah penumpang kereta dari tahun ke tahun. Pada 2013, penumpang kereta api jarak pendek maupun jarak panjang, mencapai 222 juta penumpang. Jumlah ini meningkat 26,28% pada 2014 menjadi 280,35 juta penumpang. Pada 2016, tercatat ada 352,31 juta penumpang kereta api, atau meningkat 7,7 persen dibandingkan tahun sebelumnya [3].

Banyaknya masyarakat yang memilih kereta api sebagai moda transportasi meningkatkan urgensi penjaminan keselamatan dalam penggunaan kereta api dari pihak para penyedia layanan yang terlibat. Namun, berdasarkan data investigasi kecelakaan perkeretaapian tahun 2010-2016 yang dikeluarkan oleh Komite Nasional Keselamatan Transportasi (KNKT), sebagian besar penyebab kecelakaan masih didominasi oleh kesalahan dalam faktor prasarana, khususnya patahnya rel di lokasi-lokasi terkait [4].

Pada investigasi yang dilakukan oleh KNKT terkait kecelakaan yang menjadi *highlight* pada tahun 2010-2016, identifikasi mengenai detail kondisi pengelolaan organisasi dan manajemen yang ikut menjadi penyebab terjadinya kejadian patah rel tersebut telah dilakukan. Beberapa kondisi dari hasil identifikasi yang menjadi fokus dalam penelitian ini adalah temuan sebagai berikut.

1. Tidak dilaksanakannya Keputusan Direksi PT.KAI (Persero) tahun 2013 berupa pelaporan risiko keselamatan dalam bentuk profil risiko dan Keputusan Direksi PT.KAI (Persero) tahun 2015 berupa pelaporan *Level of Safety* secara konsisten sehingga tidak terdeteksinya kondisi prasarana yang berisiko tinggi terhadap keselamatan dan menjadi prioritas utama dalam tindakan perawatan.
2. Tidak dijelaskannya standar keandalan dari perawatan berdasarkan kelas jalur kereta api sehingga tidak ada acuan/target dalam mempertahankan konsistensi hasil perawatan.
3. Belum dilaksanakannya uji berkala dari jalur kereta api di wilayah III.2.10 Resort Peninjauan, Sub Divre III.2 (Divre IV) Tanjungkarang.

Model *machine learning* dalam melakukan prediksi analisis patah rel yang dikembangkan dalam penelitian ini diharapkan mampu membantu penyelesaian masalah di industri perkereta apian dalam bidang prasarana khususnya rel patah dengan memberikan pandangan baru dalam metode penyelesaian masalah-masalah terkait organisasi dan manajemen yang menjadi fokus penelitian dengan memanfaatkan teknologi berupa *machine learning*.

II. PENELITIAN TERKAIT

Penelitian ini dilakukan berdasarkan beberapa referensi utama berupa penelitian terkait model prediksi lokasi dan patah rel kereta serta penerapan *machine learning* yang telah dilakukan sebelumnya. Berikut adalah penelitian-penelitian yang dimaksud.

A. Multivariate Statistical Model for Predicting Occurrence and Location of Broken Rails

Penelitian ini dilakukan oleh C. Tyler Dick, Christopher P.L. Barkan, Edward R. Chapman dan Mark P. Stenly pada tahun 2003. Penelitian ini dilakukan untuk meningkatkan kualitas pelayanan pengadaan jasa transportasi kereta di Amerika Serikat dalam bidang keamanan. Model prediksi yang dikembangkan dalam penelitian ini diharapkan bisa membuat proses deteksi patah rel dapat dilakukan secara lebih efektif dengan menggunakan seluruh sumber daya yang ada.

Pengembangan model prediksi dilakukan dengan menggunakan data milik Burlington Northern Santa Fe terkait *service failure* dengan detail informasi berupa tanggal, lokasi, tipe *service failure*, serta data teknis dan operasional dari lokasi setiap kejadian selama dua tahun sebelumnya. *Service failure* adalah kejadian rel patah yang teridentifikasi oleh *track circuit system* atau *track inspector* [5]. Penggunaan data *service failure* dalam pengembangan model menjadikan penelitian ini tepat untuk dijadikan referensi dalam penelitian yang sedang dilakukan. Pengembangan model dalam penelitian ini dilakukan menggunakan Statistical Analysis System (SAS) dan prosedur LOGISTIC.

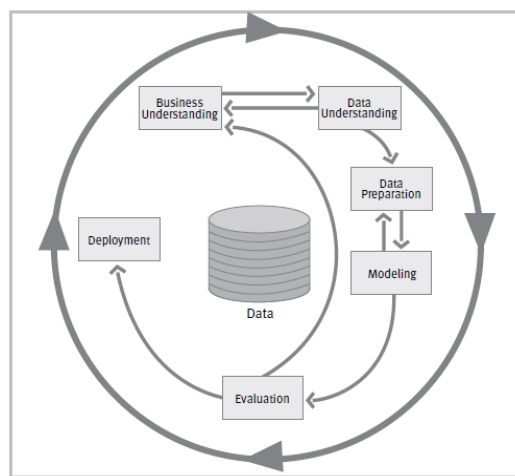
B. A Prediction Model for Broken Rails and an Analysis of Their Economic Impact

Penelitian ini dilakukan oleh Darwin H. Schafer II dan Christopher P. L Barkan pada tahun 2008. Penelitian ini merupakan penelitian lanjutan dari penelitian sebelumnya yang berjudul *Multivariate Statistical Model for Predicting Occurrence and Location of Broken Rails*. *Objective* penelitian ini adalah untuk mengembangkan model yang bisa digunakan untuk mengidentifikasi lokasi di jaringan rel yang memiliki kemungkinan terjadinya patah rel paling tinggi berdasarkan data *service failure* dan faktor-faktor lain yang kemungkinan berpengaruh. Penggunaan data *service failure* dalam pengembangan model menjadikan penelitian ini tepat untuk dijadikan referensi dalam penelitian yang sedang dilakukan. Pemodelan dalam penelitian ini dilakukan dengan menggunakan Statistical Analysis Software (SAS) dan prosedur LOGISTIC.

Sebelum penelitian dilakukan, model yang telah dihasilkan dalam penelitian sebelumnya diuji kembali untuk memastikan tingkat akurasi model saat menggunakan dataset yang berbeda dan lebih baru (data tahun 2005 dan 2006). Hasil pengujian tersebut menunjukkan bahwa tingkat akurasi menurun secara drastis dari sebelumnya. Tingkat akurasi yang sebelumnya melebihi 87,4% menurun menjadi 54,8% dalam pengujian tersebut. Hasil pengujian ditindaklanjuti dengan pemeriksaan lebih lanjut terhadap dataset *service failure* yang digunakan dalam penelitian sebelumnya. Pemeriksaan ini memberikan hasil bahwa ada kemungkinan dataset sebelumnya belum mencakup semua *service failure* yang terjadi dalam jaringan rel pada periode dataset yang digunakan. Penemuan ini menghasilkan kesimpulan terkait dibutuhkannya pengembangan model statistik yang baru dan lebih *update* untuk memprediksi lokasi terjadinya *service failure*.

III. METODOLOGI PENELITIAN

Metodologi yang dipilih dalam penelitian ini adalah CRISP-DM (*Cross-Industry Standard Processing for Data Mining*). CRISP-DM adalah metodologi standar untuk melakukan *data mining* dalam industri terkait hal tersebut [6]. Metodologi ini dikembangkan pada tahun 1996 oleh Daimler Chrysler, SpSS Inc. (saat itu bernama sebagai perusahaan bernama Integral Solusion Limited), dan NCR Corporation selalu praktisi dan perusahaan yang sudah berpengalaman dalam bidang *data mining* pada masa tersebut [7].



Gambar 1 Rangkaian proses metodologi CRISP-DM (Chapman, et al., 2000)

Metodologi CRISP-DM memiliki rangkaian fase sebagai berikut [7].

1) *Business understanding*

Fase pertama ini berfokus pada kegiatan memahami *objective* pelaksanaan *data mining* dan seluruh kebutuhan terkait yang ditinjau dari sudut pandang bisnis. Informasi ini kemudian diproses menjadi definisi masalah dalam kegiatan *data mining* dan rencana awal yang dibuat untuk mencapai tujuan yang telah diidentifikasi.

2) *Data understanding*

Fase ini dimulai dengan proses pengumpulan data awal dan melanjutkannya dengan memahami data yang tersedia, mengidentifikasi masalah-masalah terkait kualitas data, dan mendeteksi subset yang menarik untuk dijadikan hipotesis terkait informasi yang tersembunyi.

3) *Data preparation*

Fase ini mencakup seluruh aktivitas yang dibutuhkan untuk membangun dataset final yang akan digunakan dalam proses pemodelan. Kegiatan yang dilakukan dalam fase ini biasanya dilakukan secara berulang kali tanpa urutan yang kaku. Kegiatan tersebut meliputi pembuatan tabel dan rekaman, pemilihan atribut, serta transformasi dan pembersihan data untuk alat pemodelan.

4) *Modeling*

Pada fase ini dilakukan pemilihan dan penerapan berbagai teknik pemodelan data dengan parameter yang sudah dikalibrasi dengan konfigurasi yang dapat membuat hasil pemodelan berada pada tingkat akurasi paling optimal. Biasanya, setiap tipe permasalahan data mining dapat diselesaikan oleh beberapa teknik pemodelan. Penerapan beberapa teknik pemodelan juga memiliki syarat-syarat terkait kondisi data yang perlu dipenuhi. Hal ini menyebabkan perlunya kembali ke tahap pemodelan ketika fase-fase berikutnya sedang berjalan.

5) *Evaluation*

Pada fase ini dilakukan evaluasi atas model-model yang telah dibuat. Evaluasi dilakukan untuk memastikan bahwa ada model yang telah memenuhi *business objective* yang telah ditentukan di awal. Pada akhir fase ini, model final yang akan digunakan dalam *data mining* ditentukan berdasarkan hasil evaluasi tersebut.

6) *Deployment*

Pada fase ini dilakukan penerapan model terhadap media yang dapat memungkinkan pengguna memahami cara menggunakan model yang telah dibuat. Bentuk kegiatan *deployment* didasarkan kembali pada daftar kebutuhan yang telah ditentukan pada fase awal. Bentuk kegiatan *deployment* bisa dilakukan dalam berbagai bentuk mulai dari sebatas laporan, hingga penerapan sistem untuk interaksi langsung antara sistem dan pengguna. Fase *deployment* sendiri tidak termasuk dalam lingkup dari penelitian ini.

IV. HASIL DAN PEMBAHASAN

A. *Business Understanding*

Pada tahapan ini penelitian dilakukan dengan identifikasi permasalahan dan kebutuhan terkait perkereta api Indonesia, khususnya dalam bidang keamanan melalui studi pustaka. Sumber utama yang digunakan dalam tahap ini berupa laporan *media release* dari Komite Nasional Keselamatan Transportasi (KNKT) Indonesia pada tahun 2016 yang berjudul Data Investigasi Kecelakaan Perkeretaapian tahun 2010 – 2016. Laporan ini meliputi *overview* hasil investigasi kecelakaan dalam kereta api di Jawa dan Sumatera serta deskripsi secara lebih rinci mengenai *highlight*

kecelakaan pada periode tersebut sebagai contoh. Sumber-sumber lainnya yang mendukung dalam kegiatan ini adalah berita-berita dari portal berita *online*.

Pada tahun 2010-2016, terjadi 35 kecelakaan perkeretaapian di Indonesia yang diinvestigasi oleh KNKT. Jenis kecelakaan yang paling banyak terjadi adalah berjenis anjlokan/ terguling dengan jumlah sebanyak 24 dari 35 kejadian. Sembilan kejadian lainnya merupakan kecelakaan dengan jenis tumbukan antar kereta api, sementara dua kecelakaan lainnya dikategorikan sebagai jenis lain-lain.

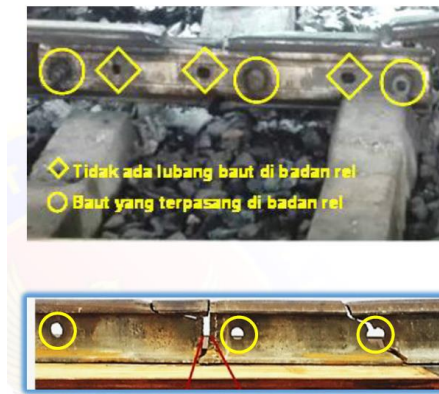
Berdasarkan hasil investigasi KNKT secara umum, faktor penyebab utama kecelakaan tersebut sebagian besar didominasi oleh faktor prasarana yang mengambil persentase sebanyak 41 % dari seluruh kejadian. Faktor penyebab utama kecelakaan terbesar selanjutnya ditempati oleh sumber daya manusia sebanyak 33 %. Sementara itu, faktor – faktor lainnya seperti sarana, operasional, dan eksternal tidak memiliki persentase yang cukup besar dengan detail berupa 19 %, 7 %, dan 0 % untuk masing-masing faktor. Dengan mempertimbangan bahwa faktor prasarana memiliki persentase terbesar sebagai penyebab utama kecelakaan kereta api, faktor prasarana dipilih sebagai lingkup dari masalah yang akan diselesaikan dalam penelitian.



Gambar 2 Anjlokan KA 3008 Km 262+100/200 Petak Jalan Antara St. Lubukrukam – St. Peninjawan, Sub Divre III.2 (Divre IV) Tanjung Karang, 1 Maret 2016 (KNKT, 2016)

Dalam laporan secara lebih medetail terkait *highlight* investigasi kecelakaan perkeretaapian tahun 2016, terdapat rincian mengenai faktor penyebab kecelakaan secara lebih deskriptif dan detail. Kecelakaan yang menjadi *highlight* pada periode tersebut adalah kecelakaan berjenis anjlokan yang terjadi pada KA 3008 di KM 262+100/200 dalam perjalanan dari St. Lubukrukam menuju St. Peninjawan, Tanjung karang pada 1 Maret 2016. Hasil investigasi menunjukkan bahwa penyebab utama dari kecelakaan tersebut adalah faktor prasarana berupa terjadinya patah (patah total dan patah gompal) dengan penyebab sebagai berikut [4].

1. Pelubangan baut pelat sambung yang tidak sesuai sehingga terbentuk awal retakan (*crack initiation*) pada tepi lubang kasar pada *web rail*, penjaralan retakan (*crack propagation*) hingga patah akhir (*total disintegration*)
2. Benturan yang berulang antara plat sambung dan kepala rel selama dilewati KA. Benturan ini disebabkan oleh penggunaan jumlah baut yang tidak sesuai dengan persyaratan teknis.



Gambar 3 Penyebab utama Anjlokkan KA 3008 (KNKT,2016)

Penemuan yang didapatkan dari investigasi kecelakaan KA 3008 dalam bidang prasarana secara lengkap adalah sebagai berikut [4] :

1. Ditemukan adanya kepala rel gompal.
2. Adanya rel patah pada sambungan rel di lokasi kejadian.
3. Adanya porositas di sambungan rel pada daerah lasan (pengelasan tidak sempurna). Hal ini disebabkan oleh metode pengelasan yang tidak sesuai prosedur dan penggunaan material yang disyaratkan.
4. Permukaan lubang baut tidak sempurna (terdapat sudut tajam pada bentuk lubang).
5. Jumlah baut di pelat sambung tidak lengkap (3 baut yang terpasang, berdasarkan spek seharusnya 6 baut).
6. Kondisi jalan rel disekitar lokasi kejadian kurang batu ballast.
7. Batalan beton pecah dan masih digunakan.
8. Secara umum kerusakan jalur KA di Divre IV tanjungkarang pada tahun 2016 adalah sebagai berikut:
 - a. Rel yang mengalami kerusakan sebanyak 54% dari total keseluruhan panjang jalur KA.
 - b. Program perawatan yang berjalan hanya dapat melingkupi 22% dari total keseluruhan jalur KA yang rusak.
 - c. Masih terdapat backlog yang tersisa sebanyak 42% dari total keseluruhan panjang jalur KA.

Hasil investigasi ini menunjukkan bahwa patah rel merupakan salah satu faktor prasarana yang mampu menyebabkan terjadinya kecelakaan kereta api yang cukup berbahaya hingga menarik perhatian untuk menjadi *highlight* kecelakaan pada tahun 2016. Dengan mempertimbangkan hal tersebut, dapat disimpulkan bahwa perawatan rel dengan kualitas yang baik dan benar dapat berperan dalam mengurangi kecelakaan kereta api yang berbahaya. Penemuan dalam bidang prasarana kemudian membawa investigasi ke beberapa penemuan dalam bidang organisasi dan manajemen yang terkait dengan manajemen perawatan rel sebagai berikut [4]:

1. Belum disusunnya standar kerusakan dan prosedur inspeksi jalan rel sebagai dasar untuk menentukan kondisi dan klasifikasi kerusakan jalan rel yang menjadi acuan dalam menentukan risiko keselamatan dan prioritas perawatan.
2. Belum dilaksanakannya uji berkala dari jalur kereta api di wilayah III.2.10 Resort Peninjawan, Sub Divre III.2 (Divre IV) Tanjungkarang.

Inspeksi rel di Indonesia saat ini masih dilakukan secara manual oleh petugas dengan berjalan kaki dari satu stasiun ke stasiun lain. Penemuan patah rel selanjutnya dilaporkan kepada pihak berwenang untuk ditindaklanjuti [8]. Tidak adanya standar

kerusakan dan prosedur inspeksi membuat analisis kerusakan sangat bergantung kepada pemahaman pelaku inspeksi sendiri. Dengan mempertimbangkan mekanisme inspeksi yang digunakan, hal ini menimbulkan kemungkinan besarnya pengaruh *human error* yang tidak bisa diidentifikasi dalam proses inspeksi. Mekanisme inspeksi yang sangat bergantung kepada sumber daya manusia juga memiliki celah untuk tidak terlaksananya uji berkala pada jalur kereta yang ditimbulkan dari *human error*.

Penelitian ini mengajukan solusi berupa otomasi proses penilai patah rel untuk meminimalisir keterlibatan manusia dalam proses tersebut. Hal ini diperlukan agar proses inspeksi yang sangat memengaruhi kondisi rel dapat dipastikan untuk selalu dilakukan dengan berdasarkan standar yang seragam.

Sistem otomasi yang diajukan dalam penelitian ini adalah berupa sistem teknologi informasi dalam bentuk aplikasi berbasis *machine learning*. Penggunaan algoritma *machine learning* dapat digunakan untuk membangun model dalam menentukan penilaian kondisi rel berdasarkan data hasil perawatan rel yang telah dilakukan di masa lampau. Penerapan algoritma *machine learning* dalam aplikasi juga memungkinkan dilakukannya pembaharuan model yang digunakan dalam prediksi seiring bertambahnya data kejadian patah rel sejak sistem digunakan. Agar perawatan preventif dapat dilakukan sehingga mengurangi kemungkinan terjadinya kecelakaan, keluaran dari sistem ini akan dibuat dalam bentuk prediksi kejadian dan lokasi patah rel. Hasil prediksi ini nantinya bisa digunakan untuk melakukan rencana perawatan dengan menggunakan sumber daya yang ada secara lebih efektif dan efisien.

Berdasarkan analisis tersebut dapat disimpulkan bahwa tujuan dari penelitian atau proses data mining yang dilakukan adalah berupa pengadaan model *machine learning* untuk mengidentifikasi area rel yang memiliki kemungkinan besar menyebabkan kejadian rel patah berdasarkan rekaman data kejadian terkait yang sudah terjadi sebelumnya dan faktor-faktor lain yang memiliki kemungkinan memengaruhi hal tersebut dalam bentuk nilai kontinu antara 0 dan 1. Nilai 0 dalam hasil prediksi menyatakan prediksi tidak adanya kejadian patah rel diarea yang diidentifikasi dan nilai 1 menyatakan sebaliknya. Lokasi yang ditinjau dalam penelitian juga dibatasi pada daerah operasi kereta api di Sumatera Selatan, khususnya DIVRE III dan DIVRE IV. Hal ini disebabkan karena kejadian rel patah paling banyak terjadi di kedua daerah operasi tersebut.

B. Data Understanding

Dengan mempertimbangkan keluaran dari hasil akhir penelitian yang berupa prediksi kejadian dan lokasi patah rel, maka data yang akan digunakan adalah data kejadian patah rel yang disertai dengan data teknis dan operasional terkait. Hal ini dilakukan juga berdasarkan pertimbangan terkait penelitian sebelumnya yang berjudul *Multivariate Statistical Model for Predicting Occurrence and Location of Broken Rails* yang menggunakan data tersebut untuk membuat model statistik prediksi kejadian dan lokasi patah rel dengan tingkat akurasi cukup tinggi pada masanya, yakni 87,4% [9]. Dataset penelitian yang akan digunakan dalam penelitian ini adalah data kejadian rel patah/gompal dan tidak patah/gompal di daerah operasi kereta api DIVRE III dan DIVRE IV pada tahun 2017 – 2019 dengan atribut berupa data kejadian dan teknis operasional sebagai berikut :

- Tanggal (hari, bulan, dan tahun)
- Lokasi (titik kilometer dihitung dari titik nol yang ditentukan)
- Kecepatan maksimal (km/jam)
- *Average tons per car* (satuan ton)

- *Average dynamic tons per car* (satuan ton)
- *Annual gross tonnage* (satuan juta ton)
- *Annual wheel passes* (satuan juta)

C. Data Preparation

Pada fase ini dilakukan persiapan data yang digunakan dalam penelitian. Fase ini dimulai dari pengambilan data dari pihak penyedia data hingga transformasi data ke bentuk yang diperlukan. Penyedia data yang digunakan dalam penelitian ini adalah bagian organisasi yang dipimpin oleh direktur pengelolaan prasarana di PT. Kereta Api Indonesia.

Data yang digunakan dalam penelitian, tidak seluruhnya diminta dari pihak penyedia data. Hal ini disebabkan ada beberapa data yang bisa diturunkan dari data lainnya. Data yang diminta dari pihak penyedia data adalah data kejadian rel patah/gompal di daerah operasi kereta api DIVRE III dan DIVRE IV pada tahun 2017– 2019 dengan atribut berupa data kejadian dan teknis operasional sebagai berikut :

- Tanggal (hari, bulan, dan tahun)
- Lokasi (titik kilometer dihitung dari titik nol yang ditentukan)
- Kecepatan maksimal kereta(km/jam)
- *Average tons per car* (satuan ton)
- *Annual gross tonnage* (satuan juta ton)

Kondisi data akhir atau data yang digunakan dalam pemodelan di penelitian ini berbentuk sebuah tabel dengan setiap barisnya mendeskripsikan satu kejadian rel patah atau gompal. Berikut adalah bentuk data yang sudah siap dipakai dalam proses pemodelan.

tahun	bulan	tangga	lokasi	ann_ton	vmax	br_cart	br_dinam	wheel	acc
2017	12	1	25,7669424	40,40513928	70	54,84245221	93,23216875	2,946997273	0

Gambar 4 Data final pemodelan

Kondisi data akhir tersebut diperoleh dengan melakukan proses transformasi dari data awal. Dalam proses transformasi tersebut data kejadian rel tidak patah dan gompal dari tahun 2017 hingga 2019 di divre III dan IV diperoleh dengan melakukan ekstraksi titik-titik lokasi antara titik awal daerah observasi (0 km) dan titik akhir daerah observasi dengan nilai yang tidak sama dengan nilai lokasi yang sudah ada pada data kejadian. Proses transformasi juga meliputi penentuan nilai-nilai atribut yang mendeskripsikan setiap data kejadian. Berikut adalah proses transformasi yang dilakukan untuk menentukan nilai-nilai atribut dari data setiap kejadian.

Tabel 1 Transformasi Data

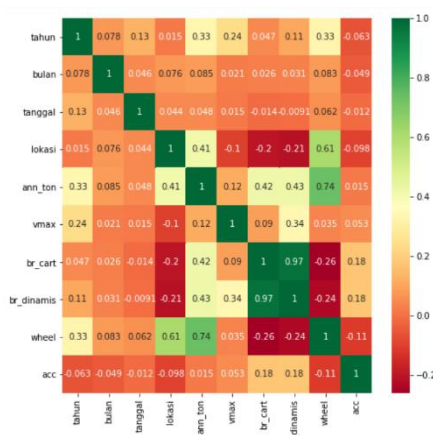
Atribut	Simbol di tabel	Proses transformasi
Tahun	Tahun	Mengambil dari data tanggal pada dokumen kejadian
Bulan	Bulan	Mengambil dari data tanggal pada dokumen kejadian
Hari	Tanggal	Mengambil dari data tanggal pada dokumen kejadian
Lokasi	Lokasi	Mengambil data kilometer dan mengubahnya menjadi data berjenis <i>float</i> dari kolom identifikasi lokasi di tabel pada dokumen kejadian yang jenisnya masih berupa string dan tercampur dengan karakter bukan angka
<i>Annual Gross Tonnage</i> (ann_ton)	ann_ton	Mengaitkan data tahun dan lokasi kejadian dengan data lokasi di tabel <i>passing tonnage</i> pada tahun terkait

Atribut	Simbol di tabel	Proses transformasi
		untuk memperoleh <i>passing tonnage</i> di lokasi terkait
Kecepatan maksimal (vmax)	vmax	Mengaitkan data tahun dan lokasi kejadian dengan data lokasi di tabel <i>passing tonnage</i> pada tahun terkait untuk memperoleh kecepatan maksimal di lokasi terkait
<i>Average tons per car</i> (br_cart)	br_cart	Mengaitkan data tahun dan lokasi kejadian dengan data lokasi di tabel <i>passing tonnage</i> pada tahun terkait untuk memperoleh kecepatan maksimal di lokasi terkait.
<i>Average dynamic tons per car</i> (br_dinamis)	br_dinamis	Melakukan perhitungan berdasarkan rumus : $A = T(1+V/100)$ dengan legenda sebagai berikut A : <i>Average dynamic tons per car</i> T : <i>Average tons per car</i> V : kecepatan maksimal kereta
<i>Annual wheel passes</i>	wheel	Melakukan perhitungan berdasarkan rumus : $W = 4T/L$ dengan legenda sebagai berikut W : <i>Annual number of wheel passes</i> T : <i>Annual gross tonnage</i> L : <i>Average tons per car</i>
Status rel patah dalam titik yang dideskripsikan (0 = tidak terjadi; 1 = ada kejadian rel patah)	Acc	Membuat kolom sendiri dan memberikan nilai 1 untuk data kejadian rel patah dan nilai 0 untuk kejadian rel tidak patah

D. Modeling

Pemodelan dilakukan menggunakan google colab dengan bahasa pemrograman python. Modul utama yang digunakan dalam pemrograman di fase pemodelan ini adalah scikit-learn. Scikit-learn adalah modul yang memungkinkan pemrograman machine learning berbasis python dan didistribusikan di bawah lisensi 3-Clause BSD [10].

Proses selanjutnya adalah analisis dan pre-processing data secara lebih lanjut. Kegiatan yang dilakukan dalam proses ini meliputi peninjauan data yang telah dimuat dengan mengecek *header* data, mengecek *outlier*, mengecek jenis data setiap atribut, mengecek *missing value* dan menanganinya, serta *feature selection*. Penanganan *missing value* dalam penelitian ini dilakukan dengan mengisi sel kosong menggunakan rata-rata dari kolom sel kosong terkait. *Feature selection* dilakukan dengan melihat keterkaitan antara atribut dalam data menggunakan *heatmap*. Atribut variabel independen (atribut selain acc) yang memiliki keterkaitan bernilai positif dengan atribut variabel dependen (acc) selanjutnya dimasukkan ke data final untuk proses pembuatan model. Hasil proses *feature selection* menunjukkan bahwa fitur-fitur yang berpengaruh pada terjadinya rel patah adalah kecepatan maksimal (km/jam), *average tons per car* (satuan ton), *average dynamic tons per car* (satuan ton), *annual gross tonnage* (satuan juta ton)



Gambar 5 Heatmap untuk feature selection

Pemodelan dalam penelitian ini akan dilakukan menggunakan berbagai macam teknik. Selanjutnya, seluruh hasil pemodelan akan dievaluasi menggunakan metode yang telah ditetapkan. Hasil evaluasi tersebut kemudian akan dibandingkan antara satu sama lain untuk menentukan model terbaik yang digunakan di sistem yang akan digunakan di kehidupan nyata. Hal ini dilakukan dengan mempertimbangkan bahwa tidak ada satu algoritma yang lebih unggul dari algoritma lain dalam menyelesaikan permasalahan semua dataset [11]. Berikut adalah teknik-teknik yang akan digunakan dalam pemodelan :

1. *Support Vector Machine* (dengan kernel='rbf')
2. *Decision Tree*
3. *Random Forest* (dengan n_estimators = 103 dan random_state = 515)
4. *K-Nearest Neighbor* (dengan n_neighbors = 19)

E. Evaluation

Sebelum pemodelan dilakukan, dataset yang ada akan dibagi dua dengan persentase sebanyak 80% untuk *training* dan 20% untuk pengujian. Setelah pemodelan dilakukan, setiap model yang dihasilkan diuji akurasi menggunakan *root mean squared error* (RMSE) dan data training yang telah dipisahkan. Model untuk deployment kemudian ditentukan dengan memilih model yang memiliki nilai RMSE terkecil dalam hasil pengujian. Eksekusi program yang telah dibuat untuk melakukan pemodelan dan evaluasinya memberikan nilai dalam matriks evaluasi sebagai berikut.

Tabel 2 Hasil evaluasi pengujian model

Metode	Nilai RMSE
<i>Support Vector Machine</i>	0.5856426400801563
<i>Decision Tree</i>	0.4806360950985497
<i>Random Forest</i>	0.47885622798803357
<i>K-Nearest Neighbour</i>	0.4862270843368316

V. KESIMPULAN

Berdasarkan hasil pengujian yang dilakukan pada proses evaluasi, dapat disimpulkan bahwa metode pemodelan *random forest* merupakan metode *machine learning* terbaik dari seluruh metode yang telah diuji untuk melakukan pemodelan dalam prediksi lokasi dan kejadian rel patah menggunakan *service failure*. Hal ini ditunjukkan dengan paling kecilnya nilai RMSE dari model yang dibentuk menggunakan algoritma *random forest*. Penelitian ini selanjutnya bisa dikembangkan

dengan melakukan *hyperparameter tuning* lanjutan dengan lebih detail pada setiap modelnya untuk mendapatkan model yang lebih baik. Dataset pemodelan juga bisa dikembangkan agar data lebih bervariasi dengan memperluas wilayah lingkup penelitian atau memperpanjang lingkup periode terjadinya data kejadian yang digunakan sehingga model bisa menjadi lebih *reliable* untuk digunakan dalam melakukan prediksi.

REFERENSI

- [1] VIVA, “Kereta Api Masih Jadi Moda Transportasi Populer,” 30 July 2017. [Online]. Available: <https://www.viva.co.id/gaya-hidup/travel/941094-kereta-api-masih-jadi-moda-transportasi-populer>.
- [2] A. Aziz, “Laba KRL Tinggi, Tapi Penumpang Masih Keluhkan Layanan Dasar,” 16 January 2017. [Online]. Available: <https://tirto.id/laba-krl-tinggi-tapi-penumpang-masih-keluhkan-layanan-dasar-cgWv>.
- [3] D. Purnamasari, “90% Pengguna Kereta Api adalah Penumpang Jarak Dekat,” 16 January 2018. [Online]. Available: <https://tirto.id/90-pengguna-kereta-api-adalah-penumpang-jarak-dekat-cDiY>.
- [4] KNKT, “Data Investigasi Kecelakaan Perkereta apian tahun 2010 – 2016,” KNKT, Jakarta, 2016.
- [5] C. T. Dick, C. P. L. Barkan, E. R. Chapman dan M. P. Stehly, “Multivariate Statistical Model for Predicting Occurence and Location of Broken Rails,” *Transportation Research Record* 1825, pp. 48-55, 2003.
- [6] IBM, “CRISP-DM Help Overview,” 2012. [Online]. Available: https://www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm.
- [7] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer dan R. Wirth, *CRISP-DM 1.0 Step-by-Step Data Mining Guide*, SpSS Inc., 2000.
- [8] L. A. Nugroho, “Begini Cara Petugas Kereta Api Awasi Jalur KA Pangrango, Susur Rel Sejauh 7 KM,” 3 January 2019. [Online]. Available: <https://bogor.tribunnews.com/2019/01/03/begini-cara-petugas-kereta-api-awasi-jalur-ka-pangrango-susur-rel-sejauh-7-km>.
- [9] D. H. Schafer dan C. P. L. Barkan, “A Prediction Model for Broken Rails and an Analysis of Their Economic Impact,” dalam *Proceedings of the American Railway Engineering and Maintenance-of-Way Association Annual Conference*, Salt Lake City, 2008.
- [10] A. Mueller, “scikit-learn 0.22.2.post1,” 2020. [Online]. Available: <https://pypi.org/project/scikit-learn/#description>.
- [11] A. Maheshwari, N. Davendralingam dan D. A. DeLaurentis, “A Comparative Study of Machine Learning Techniques for Aviation Applications,” dalam *Aviation Technology, Integration, and Operations Conference*, Atlanta, 2018.