

Penerapan Data Mining Untuk Prediksi Tingkat Kecelakaan Kerja Menggunakan Rapid Miner

Anita¹ dan Fivtatianti Hendajani²

⁽¹⁾Universitas Gunadarma,

Jl. Margonda Raya No. 100, Depok, Jawa Barat 16424

⁽²⁾STMIK Jakarta STI&K

Jl. BRI No.17, Radio Dalam, Kebayoran Baru, Jakarta Selatan 12140

{anita_sugianto, fivtatianti}@staff.gunadarma.ac.id

ABSTRAK

Keselamatan kerja berkaitan erat dengan mesin, pesawat, alat kerja dalam proses landasan tempat kerja dan lingkungannya serta cara melakukan pekerjaan. Perlindungan tersebut merupakan hak asasi yang wajib dipenuhi oleh perusahaan. Menurut perkiraan ILO (International Labour Organization), setiap tahun di seluruh dunia 2 juta orang meninggal karena masalah akibat kerja. Dari jumlah ini, 354.000 orang mengalami kecelakaan fatal. Tingkat kecelakaan fatal di negara berkembang empat kali lebih tinggi dibanding negara industri. Beberapa pekerjaan berbahaya yang mendatangkan sebagian besar kecelakaan dapat dijumpai di sektor pertanian, konstruksi, pertambangan, kehutanan dan perikanan. Paper ini bersifat deskriptif untuk mengkaji kecelakaan kerja yang terjadi berdasarkan data sekunder dan dilakukan prediksi untuk mengetahui tingkat kecelakaan kerja dengan pendekatan Data Mining menggunakan Rapid Miner. Rapid miner merupakan software pengolah data mining yang menyediakan tool untuk membuat decision tree untuk melakukan analisis terhadap data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. Data yang digunakan adalah data sekunder yang didapat dari Industrial Safety and Health Analytics Database. Konten dari basis data ini pada dasarnya mengenai catatan kecelakaan dari 12 pabrik yang berbeda di 3 negara berbeda yang setiap baris dalam data merupakan kejadian kecelakaan yang terdiri dari 439 data. Kesimpulan yang didapatkan, diantaranya secara keseluruhan dari ke-12 pabrik, ada 11 pabrik yang memiliki tingkat kecelakaan di level I; di pabrik ke 3 (level_03) memiliki tingkat kecelakaan di level IV pada Risiko Critico power lock; untuk pabrik ke 11 (local_11) tidak terdapat angka kecelakaan.

Kata Kunci : Kecelakaan kerja; Data mining; Manufaktur; Prediksi; Rapid Miner

PENDAHULUAN

Keselamatan dan Kesehatan Kerja (K3) adalah suatu pemikiran dan upaya untuk menjamin keutuhan dan kesempurnaan baik jasmaniah maupun rohani tenaga kerja pada khususnya dan manusia pada umumnya, hasil karya dan budaya untuk menuju masyarakat adil dan makmur[1]. Keselamatan kerja yaitu suatu upaya pencegahan terjadinya kecelakaan kerja. Keselamatan kerja pada dasarnya mencari dan mengungkapkan kelemahan operasional yang memungkinkan terjadinya kecelakaan. Perkiraan ILO (International Labour Organization), setiap tahun di seluruh dunia 2 juta orang meninggal karena masalah akibat kerja. Dari jumlah ini, 354.000 orang mengalami kecelakaan fatal. Disamping itu, setiap tahun ada 270 juta pekerja yang mengalami kecelakaan akibat kerja dan 160 juta yang terkena penyakit akibat kerja. Biaya yang harus dikeluarkan untuk bahaya akibat kerja ini amat besar. ILO

memperkirakan kerugian yang dialami sebagai akibat kecelakaan dan penyakit akibat kerja setiap tahun lebih dari US\$ 1.25 triliun atau sama dengan 4% dari Produk Domestik Bruto (GDP).

Tingkat kecelakaan fatal di negara berkembang empat kali lebih tinggi dibanding negara industri. Di negara berkembang, beberapa pekerjaan berbahaya yang mendatangkan sebagian besar kecelakaan kerja atau penyakit akibat kerja dapat dijumpai di sektor pertanian, konstruksi, pertambangan, kehutanan dan perikanan. Kelima sektor inilah sering didapati industri dengan tingkat resiko bahaya kerja yang paling tinggi.

Industri sektor pertambangan memiliki resiko tinggi, misalnya di pertambangan minyak dan gas bumi. Banyaknya kecelakaan yang terjadi di sektor migas seperti kebakaran, peledakan, pencemaran lingkungan dan lainnya menyebabkan industri migas memiliki potensi bahaya yang

tinggi terhadap kejadian kecelakaan kerja [3].

Paper ini bersifat deskriptif untuk mengkaji kecelakaan kerja yang terjadi berdasarkan data sekunder dan dilakukan prediksi untuk mengetahui tingkat kecelakaan dengan pendekatan *Data Mining* menggunakan *Rapid Miner*.

METODE PENELITIAN

Ditinjau dari sudut keilmuan, kesehatan dan keselamatan kerja adalah ilmu pengetahuan dan penerapannya dalam usaha mencegah kemungkinan terjadinya kecelakaan dan penyakit akibat kerja di tempat kerja [4].

Kondisi fisiologis-fiskal meliputi penyakit-penyakit kecelakaan kerja seperti kehilangan nyawa atau anggota badan, cedera yang diakibatkan gerakan yang berulang, sakit punggung, sindrom karpaltunnel, penyakit-penyakit kardiovaskular, berbagai jenis kanker seperti kanker paru-paru dan leukemia, emphysema serta arthritis. Kondisi-kondisi lain yang diketahui sebagai akibat dari tidak sehatnya lingkungan pekerjaan meliputi penyakit paru-paru putih, penyakit paru-paru coklat, penyakit paru-paru hitam, kemandulan, kerusakan system saraf pusat dan bronchitis kronis.

Kondisi-kondisi fisiologis diakibatkan oleh stress pekerjaan dan kehidupan kerja yang berkualitas rendah. Hal ini meliputi ketidakpuasan, sikap apatis, penarikan diri, penonjolan diri, pandangan sempit, menjadi pelupa, kebingungan terhadap peran dan kewajiban, tidak mempercayai orang lain, bimbang dalam mengambil keputusan, kurang perhatian, mudah marah selalu menunda pekerjaan dan kecenderungan untuk mudah putus asa terhadap hal-hal yang remeh.

Tujuan keselamatan dan kesehatan kerja, diantaranya:

- Mencegah terjadinya kecelakaan kerja
- Mencegah timbulnya penyakit akibat suatu pekerjaan.
- Mencegah / mengurangi kematian.
- Mencegah/mengurangi cacat tetap.
- Mengamankan material, konstruksi, pemakaian, pemeliharaan bangunan,

alat-alat kerja, mesin-mesin, instalasi dan lain sebagainya.

- Meningkatkan produktivitas kerja tanpa memeras tenaga kerja dan menjamin kehidupan produktifnya.
- Mencegah pemborosan tenaga kerja, modal, alat dan sumber-sumber produksi lainnya.
- Menjamin tempat kerja yang sehat, bersih, nyaman dan aman sehingga dapat menimbulkan kegembiraan semangat kerja.
- Memperlancar, meningkatkan dan mengamankan produksi industri serta pembangunan.

Data yang digunakan adalah data sekunder yang didapat dari *Industrial Safety and Health Analytics Database*. Konteks dari *database* ini untuk memberikan basis data dari pabrik manufaktur di dunia nyata yang berasal dari salah satu industri terbesar di Brasil dan di Dunia. Alasan utama IHM Stefanini berbagi data kecelakaan kerja industri ini dikarenakan kebutuhan mendesak bagi perusahaan untuk memahami mengapa karyawan masih menderita beberapa cedera/kecelakaan di pabrik. Bahkan ada yang sampai meninggal dunia [5].

Konten dari basis data ini pada dasarnya mengenai catatan kecelakaan dari 12 pabrik yang berbeda di 3 negara berbeda yang setiap baris dalam data merupakan kejadian kecelakaan. Terdiri dai 439 data dan kolom. Bentuknya seperti terlihat pada Tabel 1. Deskripsi kolom terdiri dari:

1. *Data*
Informasi waktu/tanggal terjadinya kecelakaan, terdiri dari tahun 2016 dan 2017.
2. *Countries*
Negara mana kecelakaan itu terjadi, datanya dianonimkan. Terdiri dari 3 negara.
3. *Local*
Kota tempat pabrik berada, datanya dianonimkan. Terdiri dari 12 pabrik.
4. *Industry Sector*
Sector industri mana yang dimiliki pabrik

Tabel 1. Data Safety

Data (1)	Countries (2)	Local (3)	Industry Sector (4)	Accident Level (5)	Potential Accident Level (6)	Genre (7)	Employee ou Terceiro (8)	Risco Critico (9)
01/01/2016 00:00	Country_01	Local_01	Mining	I	IV	Male	Third Party	Pressed
02/01/2016 00:00	Country_02	Local_02	Mining	I	IV	Male	Employee	Pressurized Systems

5. *Accident level* (Tingkat kecelakaan)
Dari tingkat I ke VI, mencatat seberapa parah kecelakaan itu (I berarti tidak parah tetapi VI berarti sangat parah). Tabel 2 menunjukkan tingkatan *Severity*.
6. *Potential Accident Level* (Tingkat Kecelakaan Potensial)
Bergantung pada tingkat kecelakaan, basis data juga mencatat seberapa parah kecelakaan itu (karena faktor-faktor lain yang terlibat dalam kecelakaan itu).
7. *Genre*
Apakah orang itu berjenis kelamin laki-laki atau perempuan.
8. *Employee ou Terceiro* (Karyawan atau pihak ketiga)
Jika orang yang terluka adalah karyawan atau pihak ketiga.
9. *Risco Critico* (Risiko Kritis)
Beberapa deskripsi risiko yang terlibat dalam kecelakaan.

Tabel 2. Tingkatan Severity

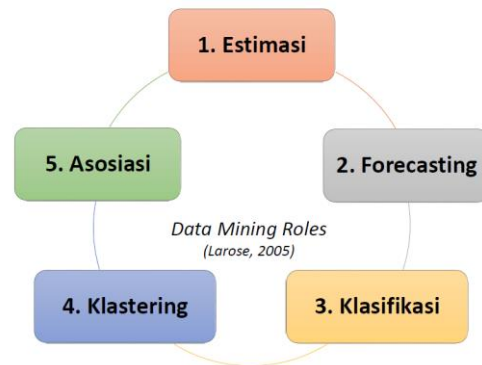
Tingkat Keparahan	Deskripsi
Level I	<i>Error</i> terjadi, namun tidak membahayakan
Level II	Membutuhkan peningkatan pengawasan, tidak perlu ada perubahan
Level III	Meningkatkan pengawasan, perubahan sementara pada bagian vital, namun tidak membahayakan
Level IV	Meningkatkan pengawasan, perubahan menyeluruh pada bagian vital, diperlukan perawatan
Level V	Meningkatkan pengawasan dan perawatan, perubahan jangka Panjang, menimbulkan bahaya sekerat

Level VI	Menyebabkan kematian
----------	----------------------

Data Mining adalah proses yang mempekerjakan satu atau lebih Teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. *Knowledge Discovery in Databases* (KDD) adalah penerapan metode saintifik pada *data mining*.

Menurut Larose, metode data mining terbagi menjadi metode Estimasi (Regresi Linier), *Forecasting* (*Neural Network*), Klasifikasi (Algoritma C4.5), Klastering (K-Means) dan Asosiasi (*FP-Growth*) [6]. Peran utama *data mining* bisa dilihat pada Gambar 1.

Tahapan *data mining* dibagi menjadi enam bagian, yaitu :



Gambar 1. Peran Utama Data Mining

1. *Pembersihan data* (*data cleaning*)
Sebelum proses *data mining* dapat dilaksanakan, perlu dilakukan proses *cleaning* pada data yang menjadi fokus KDD. Proses *cleaning* mencakup antara lain membuang 4 duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data atau informasi lain yang relevan dan diperlukan untuk KDD seperti data atau informasi eksternal.
2. *Integrasi data* (*data integration*)
Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu *database* tetapi

juga berasal dari beberapa *database* atau file teks. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Seleksi data (*Data selection*)
Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itulah data yang sesuai untuk dianalisis yang akan diambil dari *database*. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus *market basket analysis*, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.
4. Transformasi data (*Data Transformation*)
Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi menjadi beberapa interval. Proses ini sering disebut transformasi data.
5. Proses mining
Adalah sebuah proses yang paling utama pada saat metode diterapkan untuk mencari pengetahuan tersembunyi dan berharga dari data.
6. Evaluasi pola (*pattern evaluation*)
Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari Teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

7. Presentasi pengetahuan (*knowledge presentation*)

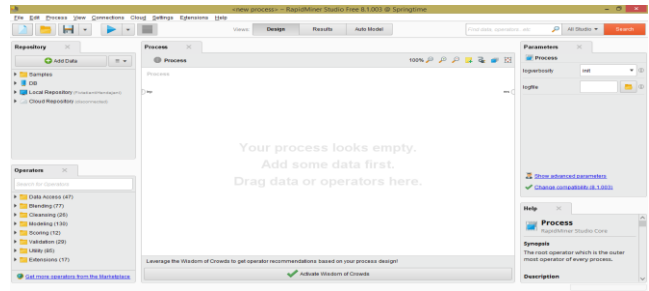
Merupakan penyajian dan visualisasi pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau 5 aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *data mining*. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil *data mining*[7].

Pohon keputusan [8] atau yang lebih dikenal dengan istilah *decision tree* ini merupakan implementasi dari sebuah sistem yang manusia kembangkan dalam mencari dan membuat keputusan untuk masalah tertentu dengan memperhitungkan berbagai macam faktor yang berkaitan dalam lingkup masalah tersebut.

Decision tree adalah struktur flowchart yang menyerupai *tree* (pohon), dimana setiap simpul internal menandakan suatu tes pada atribut, setiap cabang merepresentasikan hasil tes dan simpul daun merepresentasikan kelas atau distribusi kelas. Alur pada *decision tree* ditelusuri dari akar ke simpul daun yang memegang prediksi[9].

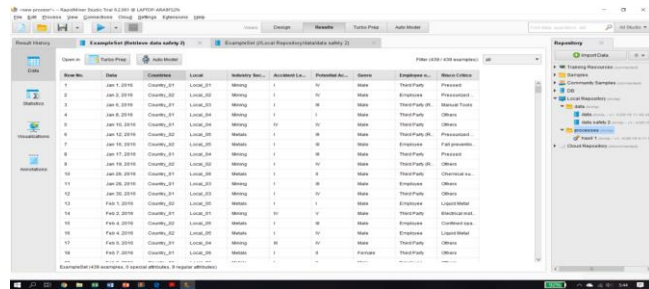
Rapid miner sebagai software pengolah data mining menyediakan tool untuk membuat *decision tree*. Hal ini tentu akan memudahkan kita membuat *decision tree* dengan menggunakan *Rapid Miner* dibandingkan membuatnya secara manual. Perangkat lunak yang bersifat terbuka (*open source*). *Rapid miner* adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan Analisa prediksi. *Rapid miner* menggunakan berbagai Teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. *Rapid miner* memiliki kurang lebih 500 operator *data mining*, termasuk operator

untuk *input*, *output*, *data preprocessing* dan visualisasi. *Rapid miner* merupakan software yang berdiri sendiri untuk Analisis data dan sebagai mesin data mining yang dapat diintegrasikan pada produknya sendiri. *Rapid miner* merupakan software pengolah data mining menyediakan tool untuk membuat *decision tree* untuk melakukan analisis terhadap data dan sebagai mesin



Gambar 3. Tampilan Halaman Muka Rapid Miner

data mining yang dapat diintegrasikan pada produknya sendiri. Data yang digunakan adalah data sekunder yang didapat dari *Industrial Safety and Health Analytics Database*.



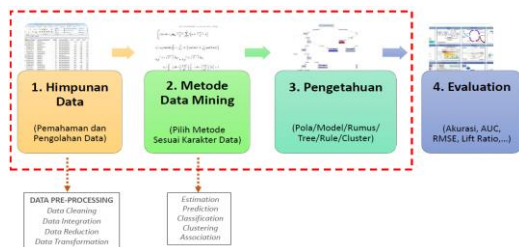
Gambar 4. Hasil Retrieve Database

HASIL DAN PEMBAHASAN

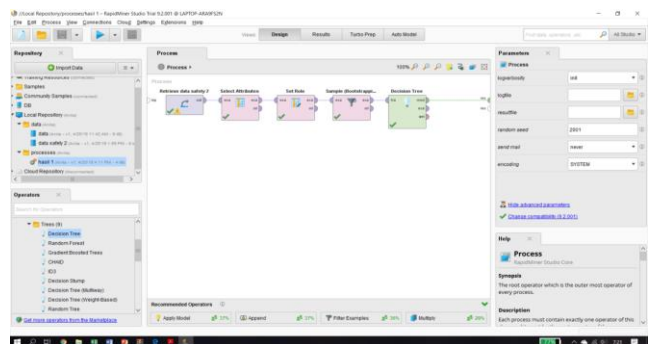
Konten dari basis data yang dibuat oleh IHM Stefanini ini pada dasarnya mengenai catatan kecelakaan dari 12 pabrik yang berbeda di 3 negara berbeda yang setiap baris dalam data merupakan kejadian kecelakaan yang terdiri dari 439 data.

Proses *data mining* dimulai dari himpunan data; metode *data mining*; pengetahuan dan *evaluation* seperti yang terlihat pada Gambar 2.

Desain untuk prediksi tingkat kecelakaan dapat dilihat pada Gambar 5. Mulai dari *Retrieve* data kemudian dilakukan *select attributes*, *set role*, *sample* dan terakhir saat *modelling* menggunakan *Decision tree*.



Gambar 2. Proses Data Mining



Gambar 5. Desain untuk Prediksi Tingkat Kecelakaan

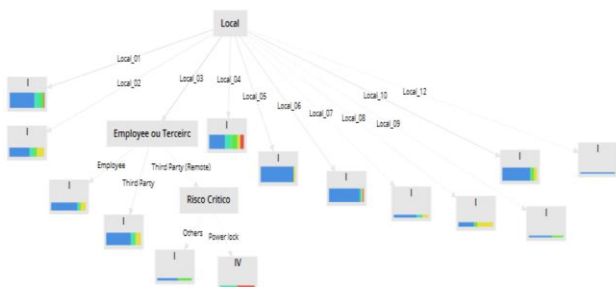
Instalasi software *rapid miner* dapat diunduh pada situs resminya yaitu www.rapid-i.com. Pada Gambar 3 menunjukkan tampilan halaman muka Rapid Miner. Dan gambar 4 menunjukkan hasil *retrieve database*. Data sekunder yang digunakan berasal dari *Industrial Safety and Health Analytics Database* yang terdiri dari 439 data.

Pada gambar 6 menunjukkan hasil *decision tree*. Dari masing2 *local* terlihat berapa besarnya tingkat kecelakaan yang terjadi mulai dari tingkat I sampai dengan tingkat VI(I berarti tidak parah tetapi VI berarti sangat parah).

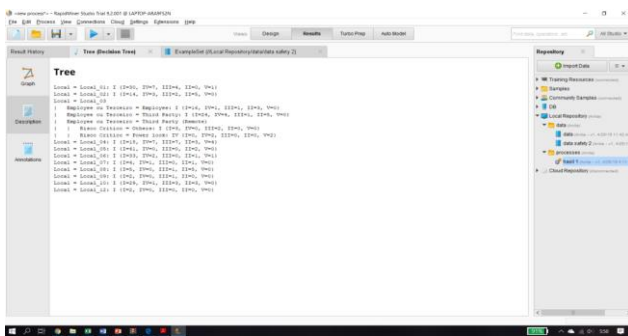
Hasil yang didapatkan dari prediksi untuk mengetahui tingkat kecelakaan kerja berdasarkan data sekunder dengan pendekatan *data mining* menggunakan *rapid miner*, diantaranya:

- Secara keseluruhan dari ke 12 pabrik, diantaranya 11 pabrik memiliki rata-rata tingkat kecelakaan di level 1 yang berarti *error* terjadi, namun tidak membahayakan.
- Pada pabrik ke-3 (*local_03*) memiliki tingkat kecelakaan di level IV yang berarti perlu meningkatkan pengawasan, perubahan menyeluruh pada bagian vital, diperlukan perawatan khususnya pada bagian *Risco Critio power lock*.
- Untuk pabrik ke-11 (*local_11*) tidak terdapat angka kecelakaan.

Selain dalam bentuk pohon, hasilnya dapat berupa *description*, seperti yang terlihat pada gambar 7.



Gambar 6. Hasil Decision Tree



Gambar 7. Description Tree

PENUTUP

Kesimpulan yang didapatkan dari prediksi untuk mengetahui tingkat kecelakaan kerja dengan pendekatan *data mining* menggunakan *rapid miner*, diantaranya:

- Secara keseluruhan dari ke 12 pabrik, diantaranya 11 pabrik memiliki rata-rata tingkat kecelakaan di level 1 yang berarti *error* terjadi, namun tidak membahayakan.

- Pada pabrik ke-3 (*local_03*) memiliki tingkat kecelakaan di level IV yang berarti perlu meningkatkan pengawasan, perubahan menyeluruh pada bagian vital, diperlukan perawatan khususnya pada bagian *Risco Critio power lock*.
- Untuk pabrik ke-11 (*local_11*) tidak terdapat angka kecelakaan.

DAFTAR PUSTAKA

- [1] Anwar Prabu Mangkunegara. 2005. Evaluasi Kinerja SDM. Refika Aditama. Bandung.
- [2] https://www.ilo.org/wcmsp5/groups/public/@asia/@ro-bangkok/@ilo-jakarta/documents/publication/wcms_120561.pdf
- [3] Ramli, S., 2010. Manajemen Risiko Dalam Perspektif K3 OHS Risk Management, Jakarta: Dian Rakyat.
- [4] Husni, Lalu. 2003. Hukum Ketenagakerjaan Indonesia. Jakarta: PT Raja Grafindo Persada.
- [5] https://www.kaggle.com/ihmstefanini/industrial-safety-and-health-analytics-database#IHMStefanini_industrial_safety_and_health_database_with_accidents_description.csv
- [6] Larose, Daniel T. 2005. Discovering Knowledge in Data An Introduction to Data Mining. John Wiley & Sons, Inc., Hoboken, New Jersey.
- [7] V Wiratna Sujarweni, Metodologi Penelitian. Yogyakarta: PUSTAKABARUPERSS,2014.
- [8] Aprilla, Dennis, dkk. 2013. Belajar Data Mining Dengan Rapid Miner. Jakarta.
- [9] Han, J & Kamber, M. 2006. Data Mining Concept and Techniques. Second Edition. Morgan Kaufmann Publishers. USA.