Seminar Pendekatan Data Mining Memprediksi Profil Sosial Masyarakat Menggunakan Aplikasi RapidMiner

Amar P. Natasuwarna

STMIK Pontianak
Jl. Merdeka Barat No. 372, (0561) 735555/ (0561) 737777
e-mail: a.pegirosa@gmail.com

Abstrak

Data mining mampu memprediksi profil sosial masyarakat. Salah satu aplikasi yang dapat digunakan untuk keperluan data mining adalah RapidMiner. Para pelajar sebagai calon pencari kerja, perlu mengenal data mining dan RapidMiner sebagai metode dan perangkat yang banyak digunakan untuk mengolah big data pada era revolusi industri 4.0. Untuk melaksanakan pengabdian kepada masyarakat, maka dilakukan seminar mengenai data mining dan RapidMiner bekerja sama dengan SMA Panca Bhakti Pontianak. Pesertanya adalah pelajar SMA Panca Bhakti Pontianak berjumlah 27 orang. Seminar ini diadakan bertujuan untuk meningkatkan pemahaman mengenai data mining dan bagaimana mengambil manfaat dengan menggunakan RapidMiner, sehingga peserta dapat menggunakan pengetahuan ini pada kegiatan pengolahan data pada masa sekarang maupun yang akan datang. Pembahasan data mining dibatasi pada metode yang paling sering digunakan yakni klasifikasi. Seminar ini merupakan kelanjutan dari seminar sebelumnya yang membahas mengenai revolusi industri 4.0, dengan contoh yang diberikan diantaranya adalah data mining dan big data. Penyampaian seminar ini dilakukan dengan cara ceramah dan tanya jawab. Hasil evaluasi menggunakan kuesioner menunjukkan bahwa terjadi peningkatan pengetahuan pelajar mengenai data mining dan RapidMiner. Dengan bekal wawasan yang telah diperoleh pada seminar ini, peserta dapat mempraktekkan big data dengan menggunakan perangkat RapidMiner yang banyak digunakan secara luas di kalangan akademik dan bisnis.

Kata kunci: Data Mining, Klasifikasi, RapidMiner, Big Data, Profil Sosial

Abstract

Data mining has the ability to predict people's social profiles. One of applications that can be used for data mining purposes is RapidMiner. Students as prospective job seekers require to understand data mining and RapidMiner as a tool that is widely used to process big data in revolution industrial 4.0 era. To carry out community service, a seminar on data mining and RapidMiner was conducted in collaboration with SMA Panca Bhakti Pontianak. The participants were student of SMA Panca Bhakti Pontianak. The number of participants who attended the public lecture was 27 students. This seminar was conducted to increase understanding of data mining and to introduce how to find benefit of RapidMiner, so participants can use these knowlege in current and future data processing activities. This data mining discussion focusing on the method most often used is classification. This seminar is a continuation of previous public lectures which discussed the industrial revolution 4.0, with exampe given of data mining and big data. Submission of the seminar is done by way of lectures, question and answers. The results of the evaluation using questioner indicate that threre is an increase in the ability of students knowledge relating to data mining and RapidMiner. Armed with the insight gained in this seminar, participants can practice big data by using the RapidMiner tool which is widely used in academic and business circles.

Keywords: Data Mining, Classification, RapidMiner, Big Data, Social Profile

1. PENDAHULUAN

Pelajar sebagai calon pencari kerja atau calon yang dapat bekerja secara mandiri memerlukan wawasan mengenai cara kerja yang berkaitan dengan revolusi industri 4.0 [1]. Kegagalan pelajar menyesuaikan diri dengan era baru ini akan meningkatkan pengangguran para pemuda usia produktif pada masa yang akan datang [2]. Hal ini disebabkan para calon pekerja ini tidak dipersiapkannya dengan wawasan dan keterampilan memadai yang harus mereka kuasai [3]. Tantangan pendidikan yang terjadi pada era revolusi industri 4.0 adalah terjadinya perubahan yang drastis dari cara berpikir dan melakukan tindakan [2]. Peran data pada masa sekarang dan akan datang semakin penting [3]. Meluasnya penggunaan media sosial, di mana setiap individu tidak hanya sebagai pengguna tetapi juga berperan sebagai produsen data. Selain itu, usaha retail seperti Alfamart dan Indomart menghasilkan data yang begitu besar setiap jamnya. Perusahaan retail seperti WalMar menghasilkan data sebesar 2,5 petabytes setiap jamnya yang dihasilkan dari transaksi pelanggan [4]. Riset-riset keilmuan salah satunya adalah memprediksi profil sosial masyarakat telah menghasilkan banyak data [4]. Akibatnya, produksi data setiap saat menjadi sangat cepat, bervariasi, dan volume semakin meningkat menghasilkan istilah yang disebut dengan big data [1]. Pelajar perlu diberikan wawasan bagaimana memanfaatkan big data tersebut. Untuk membuat big data menjadi berguna, diperlukan suatu cara kerja untuk melakukan pengolahannya yang disebut dengan data mining [1]. Supaya pelajar mendapat gambaran pekerjaan yang lebih konkret, maka diperkenalkan RapidMiner sebagai perangkat lunak yang dapat diandalkan untuk melakukan pengolahan data mining [5]. Pelajar yang dapat menguasai RapidMiner, berarti memiliki salah satu jenis pekerjaan yang dapat berguna menghadapi era revolusi industri 4.0. Hasil perbandingan antara kuesioner sebelum dan setelah seminar menunjukkan bahwa pelajar mendapatkan penambahan wawasan pengetahuan yang signifikan mengenai data mining dan RapidMiner.

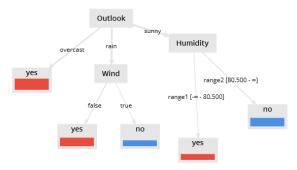
2. TINJAUAN TEORI

Data mining dapat digunakan menentukan profil dari suatu obyek dari big data yang ingin diketahui [6]. Data mining berguna untuk membantu manajemen mengambil keputusan [7]. Selain itu, hasil data mining dengan akurasi yang sangat baik dapat digunakan sebagai model untuk machine learning [6]. Berikut contoh data mining membuat model machine learning berdasarkan penelitian sosial masyarakat kota. Suatu perusahaan ojek online dari pengolahan data mining dapat menentukan suatu kota layak atau tidak apabila ojek online beroperasi pada kota tersebut. Untuk memperoleh layak atau tidak, maka dilakukan perbandingan antara model dengan data mengenai kota yang sedang dijajaki tersebut. Dengan memiliki model yang telah lengkap, perusahaan ojek online tidak perlu lagi melakukan uji coba kepada suatu kota yang sedang dijajaki.

Pemanfaatan *data mining* sangat luas [7]. *Data mining* dapat digunakan untuk seluruh keperluan manusia termasuk memprediksi profil suatu masyarakat [6]. Namun, *data mining* tidak begitu saja dapat diolah dengan data seadanya. Kebanyakan data lapangan mengalami permasalahan seperti data kotor, data terlalu besar, data tidak bisa disintegrasi, atau variasi data terlalu banyak [6]. Untuk menyelesaikan masalah-masalah tersebut, maka dilakukan terlebih dahulu yang disebut dengan *pre-processing* [6]. Ada empat hal yang dapat dilakukan pada saat *pre-processing* yakni pembersihan data, pengurangan data, integrasi data, dan transformasi data [7]. Ketika data sudah dilakukan *pre-processing*, maka data tersebut sudah siap untuk dilakukan pengolahan *data mining* [6]. Dari seluruh kegiatan *data mining*, *pre-processing* ini adalah kegiatan yang paling menyita waktu terutama apabila kita melakukan pengolahan *big data* [6]. Kegiatan *pre-processing* tidak perlu dilakukan dengan seluruh empat cara tersebut [6]. Cara-cara tersebut tergantung kebutuhan. Selain itu, kegiatan *pre-processing* tidak ada ketentuan harus dilakukan berurutan [6].

Menurut fungsinya, *data mining* terdiri dari klasifikasi, klastering, asosiasi, dan analisis *outlier* [6]. Pada seminar ini pembahasan difokuskan pada fungsi klasifikasi. Fungsi klasifikasi pada *data mining* adalah jumlah nilai label sudah diketahui [7], contohnya baik dan buruk, lulus dan gagal. Fungsi klasifikasi pada *data mining* seperti melakukan prediksi suatu obyek adalah paling sering digunakan [8]. Dengan fungsi klasifikasi, hasil-hasil prediksinya dapat digunakan untuk pengambilan keputusan. Beberapa algoritma yang mendukung fungsi klasifikasi diantaranya adalah *ID3,C45, Naive Bayes, Support Vector Machine*, dan *Neural Network* [6]. Pada seminar ini, algoritma yangdigunakan adalah *ID3*. Beberapa contoh penerapan dari algoritma *ID3* seperti identifikasi minat beli kendaraan [8], diagnosa penyakit [9], prakiraan cuaca [10], penilaian kredit [11], dan penanganan kecelakaan kerja [12].

Algoritma *ID3* menghasilkan pohon keputusan yang terdiri dari beberapa cabang [8]. Setiap akhir cabang akan menghasilkan satu label [9]. Dengan demikian, untuk setiap nilai pada suatu variabel pada akhirnya akan bermuara pada satu label yang berguna sebagai masukan untuk pengambilan keputusan [11]. Perangkat lunak *RapidMiner* mendukung algoritma *ID3* hingga visualisasi pohon keputusan [8] sebagaimana yang diperlihatkan pada Gambar 1 mengenai keputusan seseorang untuk bermain golf atau tidak berdasarkan kondisi cuaca pada saat itu.

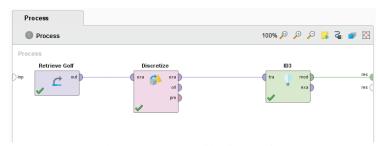


Gambar 1 Visualisasi Pohon Keputusan Menggunakan RapidMiner [6]

Pada Gambar 1 diketahui ada tiga variabel yang terdiri dari cuaca (*outlook*), kelembaban (*humidity*), dan beringin (*wind*). Dari pohon keputusan ini dapat ditentukan bahwa orang akan bermain golf atau diindikasikan dengan label *yes*, jika cuaca mendung (*overcast*), jika cuaca hujan (*rain*) dan tidak berangin (*wind*), jika cuaca cerah (*sunny*) dengan kelembaban (*humidity*) di bawah 80,5%. Sebaliknya orang tidak akan bermain golf atau label *no*, jika cuaca hujan (*rain*) dan berangin (*wind*), jika cuaca cerah (*sunny*) dan kelembaban (*humidity*) sama dengan atau di atas 80,5%. Dari Gambar 1 menunjukkan bahwa visualisasi pohon keputusan dihasilkan dengan baik oleh *RapidMiner*.

RapidMiner merupakan aplikasi yang banyak memperoleh penghargaan diantaranya pada tahun 2017, RapidMiner mendapat penghargaan dari KDNuggets sebagai the most populer general platform for data mining/data science [6]. RapidMiner pertama kali dikembangkan pada tahun 2001 oleh Raft Klinkenberg, Ingo Mierswa, dan Simon Fischer [6]. Perangkat lunak ini dapat bekerja pada lingkungan standalone dan jaringan. RapidMiner dapat berintegrasi dengan data mining, text mining, machine learning, analisis prediksi, dan analisis bisnis [6]. Untuk data mining sendiri terdiri dari banyak fungsi.

Fungsi-fungsi yang ada pada *data mining* seperti klasifikasi, klastering, asosiasi, dan analisis *outlier* dapat dilakukan menggunakan *RapidMiner* [6]. Kegiatan sebelum dilakukan pengolahan *data mining* seperti *pre-processing* tersedia pada *RapidMiner*. Pada perangkat lunak *RapidMiner*, algoritma-algoritma dan fungsi-fungsi yang dapat melakukan *pre-processing* disebut dengan nama operator [6]. Sebagai contoh operator *ID3*, operator *support vector machine* untuk algoritma, kemudian operator *filter example*, operator *join*, dan operator *replace missing values* untuk fungsi *pre-processing* [6].



Gambar 2 Proses Aplikasi RapidMiner

Gambar 2 di atas menunjukkan tampilan proses aplikasi *RapidMiner* yang terdiri dari operator komponen data, operator *pre-processing*, dan operator *data mining*. Pada proses ini operator *Retreve Golf* sebagai sumber data yang digunakan. Operator *Discretize* merupakan salah satu contoh penerapan dari *pre-processing* yang berkaitan dengan transformasi data. Operator *ID3* merupakan fungsi *data mining* klasifikasi yang dapat menghasilkan pohon Keputusan.

3. METODE PELAKSANAAN

Kegiatan seminar ini merupakan kelanjutan dari kegiatan yang sama pada semester genap tahun 2019. Seminar sebelumnya berjudul *Tantangan Menghadapi Era Revolusi Industri 4.0 – Big Data dan Data Mining*. Pembahasan seminar pada waktu itu memperkenalkan pengertian umum revolusi industri 4.0, pembahasan singkat mengenai *big data, data mining, RapidMiner*, dan *data scientist* [1]. Respon dan keingintahuan yang besar peserta pada waktu itu mengenai *data mining* dan *RapidMiner*, menginspirasi untuk melanjutkan seminar yang lebih spesifik mengenai *data mining* dan *RapidMiner*. Kegiatan seminar ini dilaksanakan pada ruang pertemuan SMA Panca Bhakti Pontianak tanggal 30 Juli 2019 dengan jumlah peserta pelajar sebanyak 27 orang yang diseleksi pihak sekolah. Waktu kegiatan dilaksanakan ketika jam sekolah. Kegiatan seminar ini berlangsung selama 90 menit. Kegiatan dimulai dengan membagikan *pre-test* dalam bentuk kuesioner selama 5 menit. Penyampaian materi seminar selama 60 menit. Materi yang diberikan terdiri dari pemahaman dasar *data mining*, manfaat *data mining*, contoh penerapan pada dikehidupan sosial masyarakat menggunakan *RapidMiner*. Penyampaian materi seminar dilakukan secara interaktif dan memberikan contoh permasalahan yang diselesaikan bersama-sama. Selanjutnya selama 20 menit dilakukan tanya jawab dan diskusi mengenai materi dan harapan pelajar mengenai kelanjutan materi ini. Menjelang berakhirnya seminar, dilakukan *post-test* dengan pertanyaan yang sama selama 5 menit.

4. HASIL PELAKSANAAN KEGIATAN

Pelaksanaan seminar data mining dan *RapidMiner* dimulai dan diakhiri dengan memberikan kuesioner. Hasil *pre-test* menunjukkan bahwa pelajar perlu mendapatkan wawasan mengenai *data mining* dan *RapidMiner*. Gambar 3 menunjukkan pelaksanaan kuesioner sebelum dimulai seminar.



Gambar 3 Kuesioner Sebelum dan Sesudah Seminar

Ada tiga pertanyaan yang sama diajukan untuk kuesioner dengan masing-masing tiga pilihan A, B, atau C sebagai mana Tabel 1 berikut ini. Ketiga pertanyaan yang sama tersebut disampaikan pada saat *pretest* dan *post-test* untuk mengukur pengetahuan mereka sebelum dan sesudah dilakukan seminar.

No	Pertanyaan	Pilih Jawaban		
1	Data mining adalah:	A. Ekstraksi informasi yang tersembunyi.		
		B. Data yang digunakan untuk transaksi bisnis.		
		C. Data yang tidak berguna.		
2	Kegunaan data mining adalah:	A. Metode untuk analisa data hingga keputusan.		
		B. Metode transaksi bisnis dan resiko.		
		C. Metode penjualan hingga keuntungan.		
3	Perangkat lunak data mining:	A. RapidMiner.		
		B. SoilMiner.		
		C. Adobe Premier.		

Tabel 1 Kuesioner

T-1-10	D1 1	TT 11	TZ	D T	1- D / T /
Tabel 2	Perbandingan	Hasıl	Kuesioner	Pre-Test	dan <i>Post-Test</i>

No	Pertanyaan	Pre-Test (%)	Post-Test (%)	
1	Data mining adalah:	3,7	62,9	
2	Kegunaan data mining adalah:	18,5	77,8	
3	Perangkat lunak data mining:	40,7	96,3	

Tabel 2 menunjukkan hasil kuesioner *pre-test* dan *post-test*. Hasil *pre-test* menunjukkan minimnya pengetahuan mereka mengenai pengertian *data mining*, kegunaan dan perangkat yang digunakan. Setelah seminar mengenai data mining dan *RapidMiner* selesai diperoleh kenaikan yang signifikan mengenai wawasan pelajar terhadap ketiga hal di atas. Dimana, terjadi peningkatan mengenai pengertian *data mining* dari 3,7% sebelum seminar menjadi 62,9% setelah seminar. Pengetahuan mengenai kegunaan *data mining* juga meningkat drastis dari 18,5% menjadi 77,8%. Selain itu pengetahuan mengenai perangkat yang digunakan untuk *data mining* meningkat dari 40,7% menjadi 96,3%.



Gambar 4 Penyampaian Materi Secara Interaktif

Supaya penyampaian materi dapat diterima dengan baik oleh para pelajar, komunikasi secara interaktif dilakukan dengan memberi contoh permasalahan sosial masyarakat dan penyelesaiannya didiskusikan bersama dengan para pelajar seperti yang ditunjukkan Gambar 4. Beberapa pelajar beraksi secara baik dan antusias dengan cara memberi tanggapan dan jawaban. Sebagian pelajar lainnya cukup berdiam diri dan bahkan ada yang sampai tidak berani menatap nara sumber.

Anak-anak Bermain Sepakbola						
Cuaca	Hujan	Berangin	Bermain			
Panas	Ya	Tidak	Tidak			
Mendung	Ya	Tidak	Ya			
Mendung	Tidak	Ya	Ya			
Panas	Tidak	Ya	Tidak			
Panas	Tidak	Tidak	?			
Mendung	Tidak	Tidak	?			

Gambar 5 Contoh Memecahkan Masalah Bersama Pelajar

Gambar 5 menunjukkan tabel yang merupakan contoh problem dan pemecahan masalah secara sederhana dengan 4 variabel dan 6 baris data tanpa perlu menggunakan *data mining*. Kemudian dari kasus ini dijelaskan bagaimana peran *data mining* terkait dengan data. Untuk data sederhana seperti pada Gambar 5, masalahnya dapat diselesaikan langsung tanpa menggunakan perhitungan dan algoritma. Namun, untuk

permasalahan dengan jumlah baris data yang besar, variabel dan variasi yang lebih banyak, manusia akan merasa kesulitan untuk menyelesaikannya.

Row No.	Survived	Age	Passenger Class	Sex	No of Siblings	No of Parent	Passenger
1	Yes	29	First	Female	0	0	211.338
2	No	2	First	Female	1	2	151.550
3	No	30	First	Male	1	2	151.550
4	No	25	First	Female	1	2	151.550
5	Yes	48	First	Male	0	0	26.550
6	Yes	63	First	Female	1	0	77.958
7	No	39	First	Male	0	0	0
8	Yes	18	First	Female	1	0	227.525
9	Yes	26	First	Female	0	0	78.850
10	Yes	80	First	Male	0	0	30
11	No	29.881	First	Male	0	0	25.925
12	No	24	First	Male	0	1	247.521
13	Yes	50	First	Female	0	1	247.521
14	Yes	32	First	Female	0	0	76.292

ExampleSet (916 examples, 1 special attribute, 6 regular attributes)

Gambar 6 Contoh Big Data

Gambar 6 menunjukkan *big data*, suatu jenis data yang sulit untuk segera dapat memecahkan masalah hanya dengan melihat secara kasat mata. Untuk *big data* ini dengan jumlah 6 jumlah atribut atau variabel dan 916 baris data diperlukan perangkat seperti *RapidMiner* dengan operator *pre-processing* dan operator algoritma *data mining* untuk mengolah data menjadi suatu model seperti pohon keputusan seperti yang telah diperlihatkan pada Gambar 1 di atas.



Gambar 7 Sesi Tanya Jawab dan Diskusi

Pada bagian akhir seminar diberikan petunjuk untuk kelanjutan belajar data mining dan RapidMiner. Pelajar yang menunjukkan perhatiannya pada data mining dan RapidMiner dapat mempelajari lebih dalam dengan melakukan instalasi perangkat lunak ini secara gratis. Untuk dapat memperoleh perangkat gratis, pelajar dapat masuk ke situs RapidMiner dengan mengetikkan rapidminer.com pada pencarian Google. Materi belajar dapat juga dipelajari pada situs RapidMiner. Pelajar dapat menemukan berbagai video sesuai dengan tema yang akan dipelajari pada situs RapidMiner tersebut. Pada saat sesi tanya jawab dan diskusi seperti pada Gambar 7, pelajar mengusulkan untuk pertemuan selanjutnya mengadakan workshop khusus mempelajari RapidMiner.

5. KESIMPULAN

Berikut ini adalah kesimpulan dari hasil pelaksanaan seminar:

- 1. Sebagai calon pencari kerja atau pekerja mandiri, pelajar perlu mendapatkan pengetahuan yang berhubungan dengan era revolusi industri 4.0. seperti *data mining* dan penerapan dengan aplikasi *RapidMiner*.
- 2. Pada seminar ini, pelajar mendapatkan tambahan wawasan *data mining* dan *RapidMiner*, caranya dengan memberi contoh permasalahan sosial masyarakat. Hal ini dapat dilihat dengan hasil perbandingan *pretest* dan *post test*. Dimana nilai *post-test* jauh lebih tinggi dari nilai yang diperoleh pada saat *pre-test*.

6. SARAN

Kelanjutan dari pembahasan lebih teknis mengenai *RapidMiner* dapat dilakukan dengan menyelenggarakan *workshop*, dimana setiap pelajar perlu dilengkapi dengan komputer dan aplikasi *RapidMiner* yang sudah diinstalasi. Sehingga, pelajar dapat lebih memahami cara penggunaan aplikasi *RapidMiner*. Khusus aplikasi *RapidMiner* yang diperuntukkan untuk para pendidik dan pelajar dapat diperoleh secara gratis pada situs resmi *RapidMiner*.

DAFTAR PUSTAKA

- [1] Natasuwarna, A.P., 2019, Tantangan Menghadapi Era Revolusi 4.0 Big Data dan Data Mining. *SINDIMAS*, Pontianak, July 29.
- [2] Yusnaini, Y., dan Slamet, S, 2019, Era Revolusi Industri 4.0: Tantangan dan Peluang dalam Upaya Meningkatkan Literasi Pendidikan. *Prosiding Seminar Nasional Program Pascasarjana Universitas PGRI Palembang*, Palembang, May 3.
- [3] Muhammad, Y., 2018, Era Industri 4.0: Tantangan dan Peluang Perkembangan Pendidikan Kejuruan Indonesia, *Sidang Terbuka Luar Biasa*, Universitas Negeri Makasar, Makasar.
- [4] Koker, D. L., 2019, Fostering collaboration amongst business intelligence, business decision makers and statisticians for the optimal use of big data in marketing strategies, *Thesis*, Department of Statistics and Population Studies, Faculty of Natural Sciences, University of the Western Cape, Cape Town.
- [5] Priyono, F., Kanti, S., Dzulfiqar, I., Amirulloh, I., Alvi, A., dan Rosiyadi, D., 2016, Analisis Sentimen Media Sosial Opini Ujian Nasional Berbasis Komputer menggunakan Metoda Naive Bayes. *Journal of Electrical and Electronics Engineering*, No.2, Vol.1.
- [6] Natasuwarna, A.P., 2019, Data Mining dengan Penerapan Aplikasi RapidMiner, Pustakaone, Jogjakarta.
- [7] Suyanto, 2017, Data Mining untuk Klasifikasi dan Klasterisasi Data, Informatika, Bandung.
- [8] Arietya, W.W., Supriyatin, W., dan Astuti, I., 2017, Penentuan Minat Konsumen dalam Membeli Mobil Menggunakan Algoritma ID3 Studi Kasus Daihatsu Wilayah Jakarta, *Jurnal KNTIA*, Vol.4.
- [9] Palengko, A.P., 2017, Sistem Pakar Diagnosa Penyakit Diabetes Melitus Menggunakan Algoritma Iterative Dichotmomiser Tree (ID3) Berbasis Android, *Doctoral Disertation*, Program Studi Teknik Informatika FTI-UKSW, Salatiga.
- [10]Purnomo, B., dan Sari, R.D.I., 2017, Penerapan Data Mining untuk Prakiraan Cuaca di Kota Malang Menggunakan Algoritma Iterative Dichotomiser Tree (ID3). *Joutica*, No.2, Vol.2.
- [11] Hayani, I., Nagni, J., dan Pasrun, Y.P., 2018, Aplikasi Data Mining untuk Penilaian Kredit Menggunakan Decision Tree Algoritma ID3 Studi Kasus PT Mandala Multi Finance Cabang Kendari, *Semantik*, No.1, Vol.4.
- [12] Puridewi, A.E., dan Nugraha, J., 2018, Perbandingan Metode Naive Bayes, Support Vector Machine dan ID3 dalam Penerapan Status Penanganan Kecelakaan Kerja. *Prosiding Sendika*, Semarang, September 10-11.