

Implementasi *Data Mining* untuk Menentukan Kesehatan Lapangan Migas dengan Metode Algoritma C4.5 (Studi Kasus PT Pertamina Hulu Indonesia)

Suastomo Widyanto
Universitas Budi Luhur, Indonesia
suastomo.widyanto@gmail.com

Abstract— *Data mining* merupakan proses untuk mendapatkan informasi yang berguna dari gudang basis data yang berupa ilmu pengetahuan. Pada penelitian kali ini, penulis akan menganalisa data dengan menggunakan algoritma C4.5 dan aplikasi Rapidminer versi 9.7 agar dapat menghasilkan suatu bentuk decision tree (pohon keputusan) sehingga harapannya perusahaan dapat menganalisa lapangan-lapangan migas yang sehat maupun yang perlu perhatian/penanganan khusus. Pada penelitian ini akan membahas dan memantau hasil angka produksi minyak dan gas berupa nilai total minyak gross, total gas gross dan total migas gross untuk menentukan suatu lapangan migas dalam menghasilkan produksi migas apakah dalam kondisi sehat/stabil atau tidak stabil/perlu penanganan khusus. Dari penelitian ini harapannya metode algoritma C4.5 dapat mengeluarkan hasil yang akurat dan mampu menganalisa tingkat kesehatan lapangan migas dalam menghasilkan produksi migasnya.

Kata kunci — *Data Mining, Algoritma C4.5, Minyak, Gas*

I. PENDAHULUAN

Data Mining adalah pengekstrakan informasi baru yang diambil dari bongkahan data besar yang membantu dalam pengambilan keputusan. Istilah data mining kadang disebut juga knowledge discovery [1]. Sedangkan menurut [2] Data Mining adalah proses ekstraksi informasi dari kumpulan data melalui penggunaan algoritma dan teknik yang melibatkan bidang ilmu statistik, mesin pembelajaran, dan sistem manajemen database. Dengan adanya data mining maka akan didapatkan suatu permata berupa pengetahuan di dalam kumpulan data – data yang banyak jumlahnya.

Hal ini sangat berguna bagi sebuah perusahaan yang bergerak di bidang tambang seperti pada penelitian ini dibahas pada sebuah perusahaan minyak dan gas ternama di Indonesia yaitu PT Pertamina Hulu Indonesia. Sebagai perusahaan holding, PHI memiliki tanggung jawab terhadap anak-anak Perusahaan untuk terus memonitor hasil produksi minyak dan gas, otomatis hal ini membuat jajaran manajemen harus menentukan strategi yang jitu dalam mengambil keputusan agar hasil angka produksi migas terus naik atau stabil.

PHI sendiri memiliki 4 anak perusahaan yaitu Pertamina Hulu Mahakam, Pertamina Hulu Kalimantan Timur, Pertamina Hulu Sanga-Sanga dan Pertamina Hulu West Ganai (data tahun 2019) yang masing-masing memiliki banyak lapangan produsen migas. Namun dari banyak lapangan tersebut, masih saja ada lapangan yang kurang stabil dalam menghasilkan migas sehingga dibutuhkan penanganan khusus namun harus dari keputusan yang tepat oleh jajaran manajemen PHI.

Pada penelitian ini, dibuat suatu aplikasi data mining dengan menggunakan algoritma C4.5 guna untuk menganalisis kemungkinan lapangan yang perlu ditangani secara khusus maupun peningkatan lapangan yang sudah stabil dalam memproduksi angka migas. Dengan dilakukannya penelitian ini diharapkan dapat memberikan hasil berupa informasi yang dapat bermanfaat bagi pihak – pihak terkait dalam melakukan pengambilan keputusan manajerial khususnya jajaran manajemen di PHI yang selama ini hanya mengandalkan data warehouse yang sudah ada namun belum dapat menghasilkan perhitungan yang tepat..

II. METODOLOGI PENELITIAN

A) Hardware dan Software

Hardware atau lebih dikenal dengan perangkat keras adalah peralatan fisik dari komputer yang dapat dilihat, dipegang, ataupun dipindahkan [13]. *Hardware* yang digunakan untuk membuat penelitian ini adalah:

- 1) Laptop dengan *processor* intel Core i7
- 2) *Memory* RAM 8 GB
- 3) SSD
- 4) *Mouse*
- 5) *Modem*

Software adalah suatu program yang berisi instruksi-intruksi yang ditulis dalam bahasa komputer yang dimengerti oleh *hardware* komputer. Kebutuhan arsitektur perangkat lunak *data mining* sebelumnya dapat digunakan untuk memperoleh kebutuhan, sehingga kebutuhan yang sudah diperoleh akan menjadi usulan untuk pengembangan aplikasi *data mining* [14]. Adapun perangkat lunak (*Software*) yang digunakan untuk membuat laporan ini adalah:

- 1) Sistem Operasi Windows 10 Enterprise
- 2) *Rapid Miner* versi 9.7

3) Microsoft Office 2016

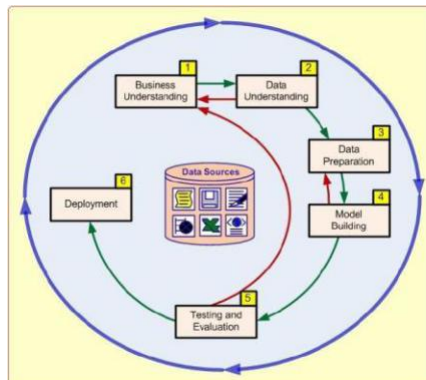
B) Analisis Sistem

Analisa sistem adalah penguraian dari sistem informasi kedalam bagian komponennya dengan maksud untuk mengidentifikasi dan mengevaluasi permasalahan-permasalahan yang terjadi dan kebutuhan yang diharapkan sehingga dapat diusulkan perbaikan-perbaikannya. Analisa sistem merupakan tahap awal dalam perancangan dan pengembangan sebuah sistem yang akan dirancang, karena tahap inilah akan diukur dan dievaluasi tentang kinerja dari sistem yang dirancang.

Dalam melakukan analisis sistem terlebih dahulu harus mengetahui dan memahami sistem, untuk menganalisa sistem diperlukan data dari sistem untuk dianalisa. Data yang diperlukan adalah hal-hal yang diperlukan untuk definisi data. Analisa data merupakan tahap untuk melakukan penganalisaan terhadap data-data yang dibutuhkan untuk perancangan sistem yang akan dibuat, dalam hal ini penulis mengambil data dari survei yang berhubungan dengan tema penelitian, untuk mencari informasi menyusun teori-teori yang berhubungan dengan dengan pembahasan sehingga terjadi perpaduan kompleks antara satu dengan yang lainnya.

Pengelompokkan dalam memprediksi kesehatan lapangan migas antara stabil atau tidaknya guna untuk menambahkan tingkat kualitas produksi migas ke depannya. Ada beberapa tahap yang dilakukan dalam melakukan eksperimen ini, penulis menggunakan model *Cross-Standard Industry For Data Mining* (CRISP-DM) [15] yang terdiri dari 6 tahap, yaitu:

1. *Business/Research Understanding Phase*
2. *Data Understanding Phase* (Fase Pemahaman Data)
3. *Data Preparation Phase* (Fase Pengolahan Data)
4. *Modeling Phase* (Fase Permodelan)
5. *Evaluation Phase* (Fase Evaluasi)
6. *Deployment Phase* (Fase Penyebaran)



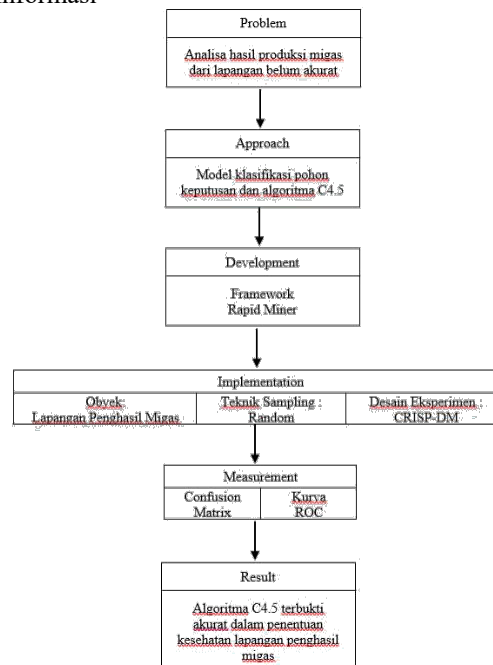
Gambar 2. Tahap CRISP-DM (*Cross Industry Standard Process for Data Mining*)

Ada beberapa tahap dalam menggunakan *Rapid Miner 9.7*:

- 1) Untuk menganalisa, dibutuhkan data training. Data training yang akan dimasukkan kedalam *Rapid Miner 9.7* bisa dalam format .csv, .xls, .mdb dan lain-lain. Data yang digunakan penulis adalah data dalam format .xlsx.
- 2) Buka program *Rapid Miner 9.7*, kemudian akan muncul tampilan awal. Untuk memasukkan data training yang telah dibuat sebelumnya, pilih menu File – Import Data – Import .xlsx file.
- 3) Tampilan jendela *Data Wizard* dengan total 5 langkah. Pada langkah ke-1 ini tentukan nama file yang berisi *data training* dalam direktori kemudian pilih *Next* 4) *Data training* yang sebelumnya disimpan, akan tersimpan otomatis ke dalam *Repositories*. Pilih tab *Repositories* – *NewLocalRepository* – *data_training*. Geser *data_training* ke area *Main Process*. Untuk menambahkan model, pilih tab *Operators-Modelling-Classification and Regression-Tree Induction-Decision Tree*. Geser *Decision Tree* ke area *Main Process* dan hubungkan.
- 4) Untuk melihat hasilnya, pilih *process* – *Run* maka akan tampil hasil berupa pohon keputusan.

C) Kerangka Pemikiran

Kerangka pemikiran dari penelitian ini, dimulai dari *problem* (permasalahan) analisa hasil produksi migas dari lapangan masih belum akurat kemudian dilanjutkan tahapan *approach* (model) yaitu algoritma C4.5 untuk memecahkan permasalahan. Untuk mengembangkan aplikasi (*development*) berdasarkan model yang dibuat, digunakan aplikasi *Rapid Miner* versi 9.7. Tahap berikutnya yaitu *implementation* (implementasi), pada tahap ini objek implementasi dilakukan pada lapangan penghasil migas, teknik *sampling* menggunakan *random sample*, dan desain eksprimennya digunakan CRISP-DM. Kerangka pemikiran dalam penelitian ini disajikan pada Gambar 3.



Gambar 3. Kerangka Pemikiran

D) Pengujian Sistem

Pengujian Sistem digunakan untuk memproses kebenaran data. Untuk pengujian data maka dibuat perancangan pengujian menggunakan aplikasi *Rapid Miner* versi 9.7 dan database DW_PHI yaitu *data warehouse* angka produksi migas yang di export ke dalam *format Excel*. Pengujian tersebut dilakukan secara langsung di PT Pertamina Hulu Indonesia. Penelitian ini menggunakan pengujian *White Box Testing*. *White Box Testing* adalah pengujian yang didasarkan pada pengecekan terhadap detail perancangan, menggunakan struktur kontrol dari desain program secara procedural untuk membagi pengujian ke dalam beberapa kasus pengujian. Secara sekilas dapat diambil kesimpulan *White Box Testing* merupakan petunjuk untuk mendapatkan program yang benar secara 100%. Pengujian *White Box* bertujuan untuk:

- 1) mengetahui cara kerja suatu perangkat lunak secara *internal*.
- 2) menjamin operasi-operasi internal sesuai dengan spesifikasi yang telah ditetapkan dengan menggunakan struktur kendali dari prosedur yang dirancang.

Pelaksanaan pengujian *White Box*:

1. Menjamim seluruh *independent path* dieksekusi paling sedikit satu kali. *Independent path* adalah jalur dalam program yang menunjukkan paling sedikit satu kumpulan proses ataupun kondisi baru.
2. Menjalani *logical decision* pada sisi dan *false*.
3. Mengeksekusi pengulangan (*looping*) dalam batas-batas yang ditentukan.
4. Menguji struktur data internal.

Berdasarkan konsep pengujian; *White box (structural) testing/glass box testing*: memeriksa kalkulasi *internal path* untuk mengidentifikasi kesalahan. Langkah-langkah white box:

Mendefinisikan semua alur logika
Membangun kasus untuk digunakan dalam pengujian
Melakukan pengujian.

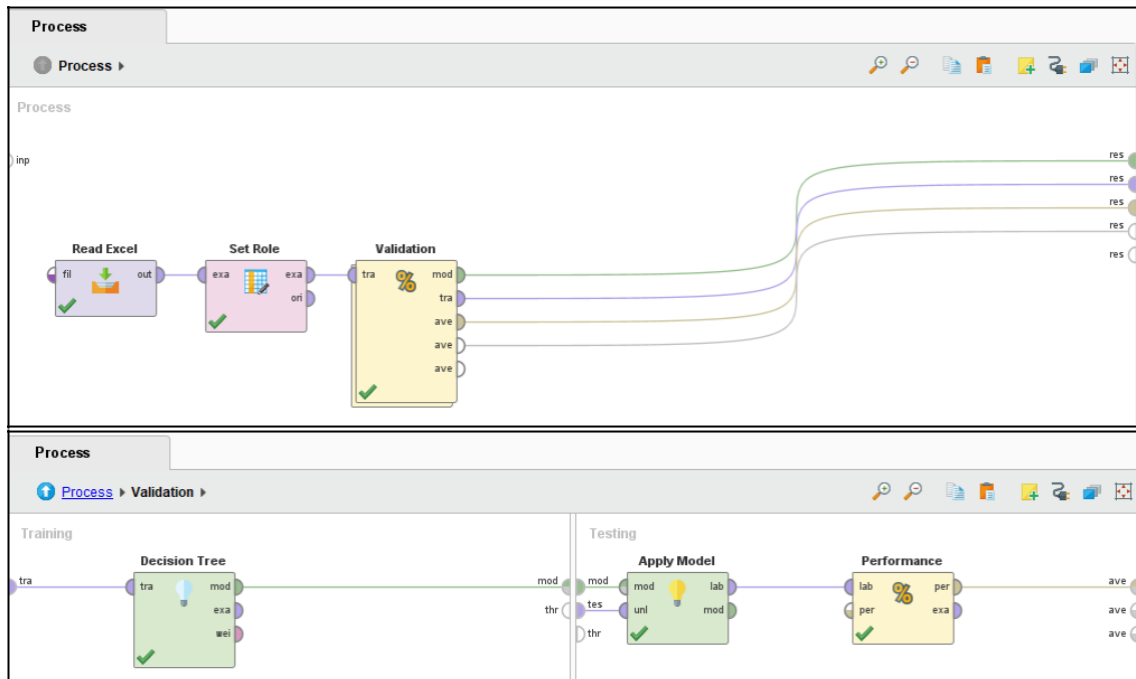
III. HASIL DAN PEMBAHASAN

A) Validasi dan Evaluasi

Adapun pada pembahasan ini, peneliti bertujuan untuk menganalisa prediksi kestabilan lapangan penghasil migas dengan menerapkan teknik klasifikasi *data mining* yaitu dengan algoritma decision tree C4.5. Pada tahap pengujian model ini data yang digunakan telah melewati tahap *pre-processing*. Desain model yang akan digunakan ditunjukkan pada Gambar 4 dan terbagi menjadi 1 desain model dan 3 hasil akhir yaitu Proses Minyak (*Oil*), Proses Gas (*Gas*) dan Proses Migas (*Oil & Gas*).

- 1) *Retreiving data*: Peneliti mengambil file dari format *.xlsx*.
- 2) *Validation*: Metode validation yang digunakan dalam penelitian ini adalah teknik *sampling*.
- 3) ID3: Metode algoritma yang digunakan.

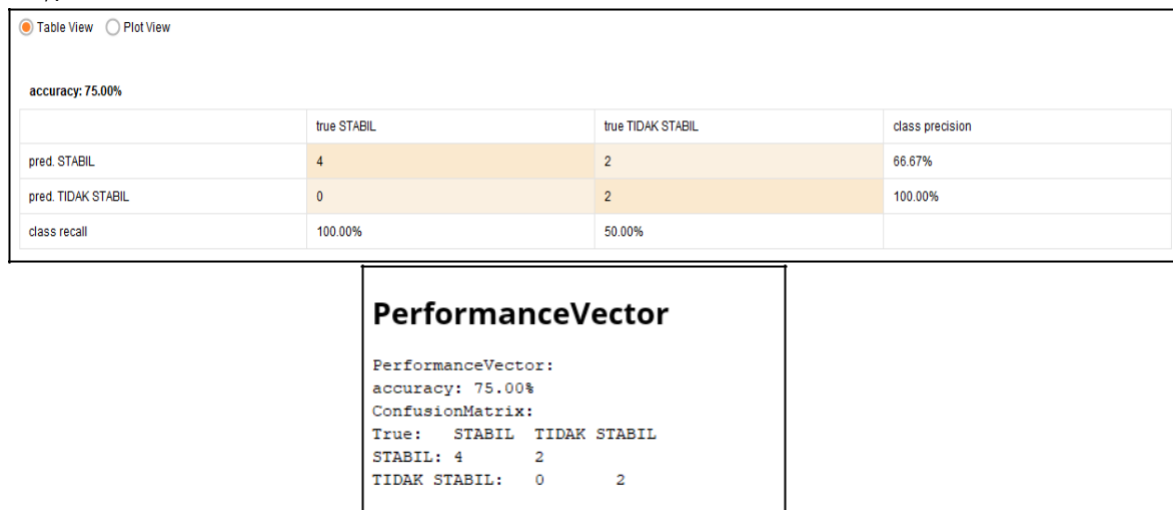
- 4) *Decision tree*: Metode klasifikasi yang digunakan dalam penelitian ini.
- 5) *Apply Model*: Operator yang digunakan dalam penelitian ini adalah C4.5.
- 6) *Performance*: Operator yang digunakan untuk mengukur performa akurasi dari model.



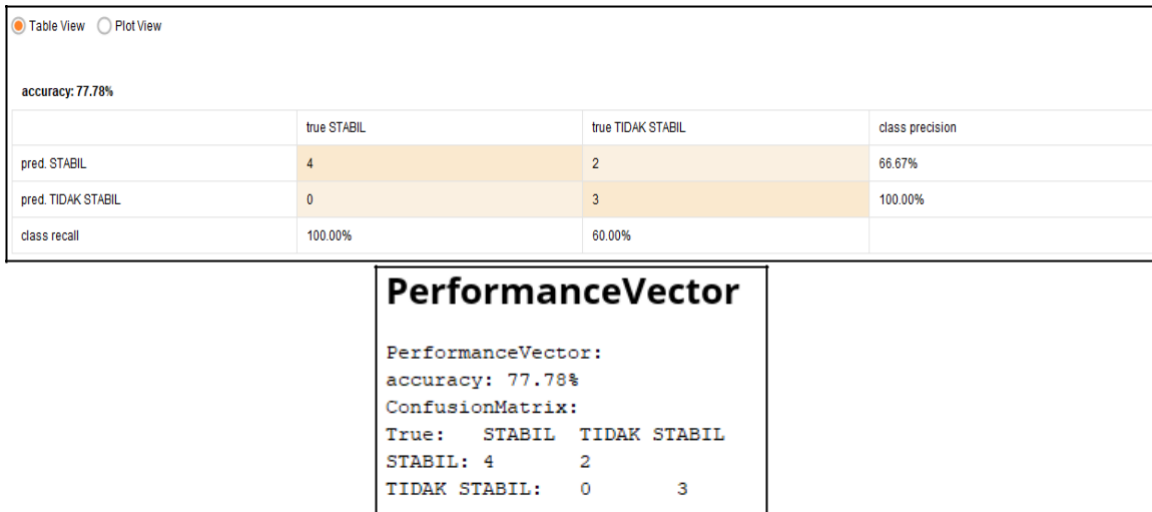
Gambar 4. Desain Model Metode C4.5

Dan dibawah ini adalah hasil akhir untuk masing-masing proses yang terlihat pada gambar 5, gambar 6 dan gambar

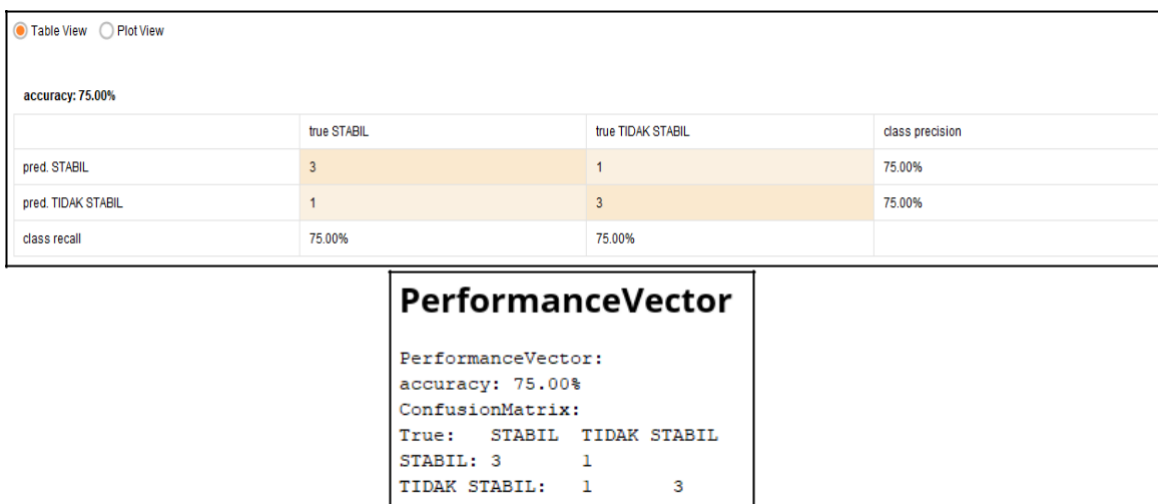
7.



Gambar 5 Hasil Proses Minyak (Oil)



Gambar 6 Hasil Proses Gas (Gas)



Gambar 7 Hasil Proses Minyak & Gas (Oil & Gas)

Pengujian dilakukan menggunakan data dari populasi *data training*. Jumlah *data training* yang akan digunakan sebanyak 28 karena memang lapangan penghasil migas di PHI hanya sebanyak 28 lapangan saja. Kemudian tingkat kesalahan yang digunakan adalah sebesar 1% baik prediksi stabil dan tidak stabil dilakukan secara acak (*simple random sampling*).

Dilihat dari hasil masing-masing metode penyeleksian atribut, hasilnya menunjukkan kesamaan. Dari hasil pengujian diatas akan dievaluasi tingkat akurasi menggunakan model yaitu menggunakan confusion matrix.

B) Evaluasi model confusion matrix

Evaluasi model disini peneliti menggunakan tabel seperti matrix di bawah ini:

	C1	C2
C1	True positives	False negatives
C2	False positives	True negatives

True positives merupakan tupel positif didata set yang diklasifikasikan positif. *True negatives* merupakan tupel negatif di data set yang diklasifikasikan negatif. *False positives* adalah tupel positif didata set yang diklasifikasikan negative sedangkan *False negatives* merupakan jumlah tupel negatif yang diklasifikasikan positif. Kemudian masukan data uji yang ada kedalam model *confusion matrix*.

Setelah data diuji dimasukan kedalam *confusion matrix*, hitung nilai-nilai yang telah dimasukan tersebut untuk dihitung jumlah *sensitivity*, *specificity*, *precision* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah *true positives* terhadap jumlah tupel yang *positives* sedangkan *specificity* adalah perbandingan jumlah *true negatives* terhadap jumlah tupel yang *negatives*. Terlihat pada gambar 5, gambar 6 dan gambar 7 di atas ini yang menunjukan nilai *accuracy*, *recall*, dan *precision* yang dihasilkan oleh Rapid Miner 9.7 menggunakan model *confusion matrix*.

C) Hasil Pengujian

Berdasarkan hasil pengujian dengan sistem yang telah dilakukan di PT Pertamina Hulu Indonesia, penulis dapat mengetahui bahwa terdapat kelebihan dan kekurangan dari sistem lama dan sistem baru tersebut.

Dengan menggunakan sistem lama, para staf masih menggunakan sistem perkiraan saja sehingga tingkat kesalahan untuk memprediksi tingkat kestabilan lapangan penghasil migas masih besar. Dengan adanya sistem baru, keabsahan dan isi data lebih baik daripada penggunaan kertas yang masih digunakan saat ini. Rata – rata waktu yang digunakan dalam satu rangkaian proses juga terbukti lebih singkat daripada sistem yang sekarang karena proses bisnis yang lebih efisien. [16]

Sedangkan pada saat penulis menggunakan teknik *data mining* ini tingkat kesalahan dalam memprediksi kesehatan lapangan migas tersebut stabil atau tidaknya dapat dikurangi dengan tingkat kesalahan 1%.

Dengan menggunakan sistem lama, pengambilan keputusannya masih kompleks dan global. Sedangkan setelah menggunakan sistem baru daerah pengambilan keputusan yang sebelumnya kompleks dan sangat global, dapat diubah menjadi lebih simple dan spesifik. Dengan menggunakan sistem lama, seorang penguji masih kesulitan dalam menganalisis untuk mengestimasi baik itu distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. Sedangkan dengan sistem baru, dalam analisis, dengan kriteria dan kelas yang jumlahnya sangat banyak, seorang penguji biasanya perlu untuk mengestimasi baik itu distribusi dimensi tinggi ataupun parameter tertentu dari distribusi kelas tersebut. Metode pohon keputusan dapat menghindari munculnya permasalahan ini dengan menggunakan kriteria yang jumlahnya lebih sedikit pada setiap node internal tanpa banyak mengurangi kualitas keputusan yang dihasilkan.

Dengan menggunakan sistem lama, membutuhkan waktu lama dalam penentuan keputusan. Sedangkan dengan sistem baru waktu yang diperlukan dalam penentuan keputusan lebih cepat dengan menggunakan teknik *data mining* tersebut.

Dengan menggunakan sistem yang baru, terlihat hanya ada beberapa lapangan penghasil migas yang bisa diprediksi akan terus stabil menghasilkan, sedangkan sisanya masuk dalam kategori perlu penanganan khusus (*Well Intervention*) sehingga ke depannya lapangan yang tidak stabil akan terus menghasilkan hasil produksi migas yang baik untuk PT Pertamina Hulu Indonesia.

IV. KESIMPULAN

Dalam penelitian ini yang berjudul “Implementasi *Data Mining* Untuk Menentukan Kesehatan Lapangan Migas. Dengan Metode Algoritma C4.5 (Studi Kasus: PT Pertamina Hulu Indonesia), penulis menerapkan algoritma C4.5 dengan menggunakan 7 (tujuh)

parameter yaitu, *Field_ID*, *Region*, *Result_Production*, *Status_Aktif*, *Total_Actual_Production* (Minyak, Gas dan Migas), *Target* (Minyak, Gas dan Migas) dan Status Lapangan. Kemudian untuk pengolahan data penulis menggunakan aplikasi pada Rapid Miner versi 9.7.

Dari hasil penelitian terbukti bahwa algoritma C4.5 lebih akurat dibandingkan analisa yang dilakukan oleh analis PHI. Hal ini dibuktikan dengan hasil evaluasi penelitian bahwa algoritma C4.5 mampu menganalisa tingkat kesehatan lapangan penghasil migas. Walaupun harapannya model algoritma C4.5 ini akan mungkin diterapkan dan mungkin berjalan dengan baik di dalam sistem, namun ada hal yang harus ditambahkan untuk menambah akurasi algoritma C4.5, yaitu:

1. Melakukan pruning terhadap algoritma C4.5 sehingga pohon yang terbentuk tidak terlalu besar bahkan mungkin untuk jumlah data yang besar sekalipun. Ini dilakukan untuk mengefisienkan kinerja dari algoritma C4.5 tanpa mengurangi keakuratannya.
2. Pada riset selanjutnya dapat digunakan metode seleksi atribut yang lain seperti *Chi-Square* untuk ketepatan penyeleksian atribut.
3. Menerapkan algoritma C4.5 kedalam data yang lebih besar untuk menguji akurasi algoritma.

DAFTAR PUSTAKA

- [1] Husain, Ahmad, (2018). Analisis Data Lifting Migas Menggunakan Metode C4.5 Pada Asosiasi Daerah Penghasil Migas, Jurnal String Vol. 2 No. 3 April 2018 - p-ISSN: 2527 – 9661 e-ISSN: 2549 - 2837.
- [2] Yanto, R., & Khoiriah, R. (2015). Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat. Creative Information Technology Journal. <https://doi.org/10.24076/citec.2015v2i2.41>
- [3] Browne, A., & Wildavsky, A. (1983). Implementation as exploration. In Implementation.
- [4] Prasetyo, Eko, (2012). Data Mining, Andi Yogyakarta, 356 Halaman.
- [5] Eska, J. (2016). Penerapan Data Mining Untuk Prekdiksi Penjualan Wallpaper Menggunakan Algoritma C4.5 STMIK Royal Ksianan. JURTEKSI (Jurnal Teknologi Dan Sistem Informasi).
- [6] Solichin, A. (2019). Comparison of Decision Tree, Naïve Bayes and K-Nearest Neighbors for Predicting Thesis Graduation. International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). <https://doi.org/10.23919/EECSI48112.2019.8977081>
- [7] Meilina, P. (2015). Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Decision Tree dan Regresi. Jurnal Teknologi. <https://doi.org/10.24853/jurtek.7.1.11-20>

- [8] Novandya, Adhika., Oktria, I. (2017). Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi. Jurnal Format.
- [9] Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. Jurnal Edik Informatika.
- [10] Kusriani, Luthfi, E. T., (2009), Algoritma Data Mining, Andi Offset, Yogyakarta.
- [11] Cynthia, E. P., & Ismanto, E. (2018). Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji. Jurasik (Jurnal Riset Sistem Informasi Dan Teknik Informatika). <https://doi.org/10.30645/jurasik.v3i0.60>
- [12] Han & Kamber, (2006). Data Mining Concept and Thenique. Morgan Kauffman. San Fransisco.
- [13] Supriyanto, (2011). Pengenalan Komputer. Graha Ilmu. Yogyakarta.
- [14] Surya, I. (2017). Kebutuhan Perangkat Lunak Untuk Aplikasi Data Mining. Jurnal Nasional Teknologi Dan Sistem Informasi. <https://doi.org/10.25077/teknosi.v3i2.2017.233-240>
- [15] Feblian, D., & Daihani, D. U. (2017). Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada PT X. Jurnal Teknik Industri. <https://doi.org/10.25105/jti.v6i1.1526>
- [16] Ansori, A. (2018). Studi Pemanfaatan Internet of Things dan Data Mining untuk Pengawasan Bahan Bakar Minyak (Studi Kasus: Perusahaan Pelayaran Penumpang Nasional). Wave: Jurnal Ilmiah Teknologi Maritim. <https://doi.org/10.29122/jurnalwave.v12i1.2915>