

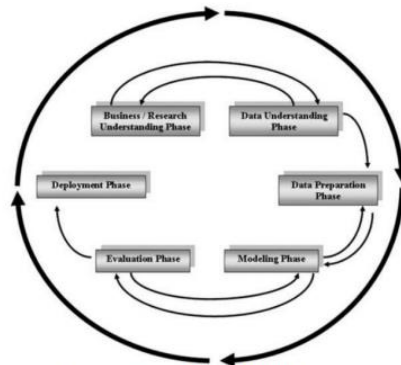
OPTIMASI ALGORITMA DATA MINING MENGGUNAKAN BACKWARD ELIMINATION UNTUK KLASIFIKASI PENYAKIT DIABETES

Diabetes merupakan penyakit kronis yang ditandai dengan tingginya kadar gula (glukosa) dalam darah yang melebihi batas normal. Menurut situs resmi World Health Organization (WHO) sekitar 422 juta orang di seluruh dunia menderita diabetes, mayoritas yang tinggal di negara berpenghasilan rendah dan menengah dan 1,5 juta kematian secara langsung dikaitkan dengan diabetes setiap tahun. Menurut situs resmi World Health Organization (WHO) sekitar 422 juta orang di seluruh dunia menderita diabetes, mayoritas yang tinggal di negara berpenghasilan rendah dan menengah dan 1,5 juta kematian secara langsung dikaitkan dengan diabetes setiap tahun. Di Indonesia sendiri kasus diabetes tidak kalah banyaknya. Dari situs data databox kita dapat lihat Indonesia berada di posisi ke-5 negara dengan kasus diabetes tertinggi setelah Amerika Serikat. Dengan banyaknya kasus diabetes di Indonesia dihasilkan data dari pasien diabetes dan dari data pasien diabetes tersebut bisa diolah menggunakan teknik data mining untuk melakukan deteksi dini penyakit diabetes.

Tujuan penelitian ini adalah untuk mengetahui algoritma terbaik dari segi nilai akurasi dan nilai AUC tertinggi untuk melakukan klasifikasi atau deteksi dini penyakit diabetes dan algoritma yang dibandingkan adalah C4.5, Naïve Bayes, dan K-Nearest Neighbor. Pengambilan data yang dilakukan adalah dengan menggunakan dataset statistic yaitu pengambilan data dari sebuah bank data. Dalam penelitian ini digunakan dataset “Early stage diabetes risk prediction dataset” yang diambil dari situs UCI Machine Learning Repository dengan 520 record dan juga 17 atribut didalamnya.

Diabetes diklasifikasikan menjadi beberapa tipe yaitu diabetes tipe 1, diabetes tipe 2, dan diabetes gestasional [10]. Diabetes tipe 1 dapat terjadi dikarenakan terdapat kerusakan sel autoimun pada pankreas yang menyebabkan produksi insulin hilang. Dan diabetes tipe 2 dapat terjadi dikarenakan hilangnya kemampuan tubuh untuk merespon insulin dan menyebabkan tubuh tidak bisa menyerap glukosa dan menyebabkan tumpukan gula dalam darah. Penyakit diabetes yang berkelanjutan dapat menyebabkan komplikasi, yang sering terjadi adalah stroke, amputasi, kebutaan, kecacatan, dan lain-lain.

Penelitian ini menggunakan framework data mining yaitu CRISP-DM atau cross industry standart process for data mining. Menurut CRISP-DM, data mining memiliki siklus hidup yang terdiri dari enam fase, dan fase tersebut bersifat adaptif, yaitu fase berikutnya sangat bergantung pada hasil yang terkait dengan fase sebelumnya. Ketergantungan paling signifikan antar fase ditunjukkan oleh panah. Keenam fase tersebut dapat dilihat pada gambar 1 berikut :



Gambar 1. Tahapan CRISP-DM

Ketergantungan paling signifikan antar fase ditunjukkan oleh panah. Berikut adalah fase CRISP-DM :
 Dalam siklus CRISP-DM terdapat 6 (enam) fase yaitu sebagai berikut:

1. Fase pemahaman bisnis (Business Understanding Phase)
 - a. Menentukan tujuan dan kebutuhan proyek secara detail dalam bisnis.
 - b. Mengubah tujuan dan batasan menjadi rumusan masalah data mining.
 - c. Menyiapkan strategi untuk mencapai tujuan.
2. Fase Pemahaman data (Data Understanding Phase).
 - a. Mengumpulkan data.
 - b. Menggunakan analisis penyelidikan data.
 - c. Mengevaluasi kualitas data.
3. Fase pengolahan data (Data Preparation Phase).
 - a. Menyiapkan data awal.
 - b. Pilih kasus dan variable yang akan dianalisis.
 - c. Jika dibutuhkan, lakukan perubahan beberapa variabel.
 - d. Siapkan data awal untuk pemodelan.
4. Fase Pemodelan (Modeling Phase)
 - a. Memilih dan implementasi model.
 - b. Jika dibutuhkan, Kembali ke proses pengolahan data untuk menjadikan data ke dalam bentuk yang sesuai dengan kebutuhan teknik data mining.
5. Fase Evaluasi (Evaluation Phase)
 - a. Evaluasi satu atau lebih model yang dilakukan pada fase pemodelan untuk mendapatkan kualitas dan efektivitas sebelum dipublish.
 - b. Menentukan apakah model memenuhi tujuan atau tidak.
 - c. Menentukan apakah masalah atau penelitian yang tidak tertangani dengan baik.
 - d. Mengambil keputusan yang berkaitan dengan penggunaan hasil dari data mining.
6. Fase Penyebaran (Deployment Phase)
 - a. Menggunakan model yang dihasilkan.
 - b. Contoh sederhana penyebaran : pembuatan laporan.

Referensi :

<https://ejournal.undiksha.ac.id/index.php/janapati/article/view/45282/pdf>