

LAPORAN
PENERAPAN ALGORITMA K-NEAREST NEIGHBOR DALAM
KLASIFIKASI PENYAKIT DIABETES
Sebagai **TUGAS UJIAN AKHIR SEMESTER KECERDASAN BUATAN**



DISUSUN OLEH:
KELOMPOK IV

| | |
|-----------------------------|------------------------|
| INTAN SUHARTINA | (20.01.013.038) |
| NATASYA AWRA FADILAH | (20.01.013.065) |
| NURUL HIDAYAT | (20.01.013.057) |
| YUNI SUKANA | (20.01.013.031) |
| WIDIYA LIS SUSANA | (211001002) |

FAKULTAS REKAYASA SISTEM
UNIVERSITAS TEKNOLOGI SUMBAWA

2021

KATA PENGANTAR

Puji syukur kehadiran Tuhan YME. yang menjadi pelindung dalam segala aktifitas penulis dan menguatkan daya semangat penulis sehingga dapat menyelesaikan laporan ini sebagai salah satu persyaratan dalam menyelesaikan Tugas Mata Kuliah Kecerdasan Buatan (Artificial Intelligent) pada Program Studi Teknik Informatika, Universitas Teknologi Sumbawa.

Meskipun dalam penyusunan laporan ini, banyak menghadapi berbagai macam hambatan, rintangan dan tantangan yang harus dilalui, tetapi berkat pertolongan Tuhan YME. dan berbagai pihak sehingga laporan ini dapat terselesaikan.

Dengan segala keterbatasan waktu dan kemampuan yang ada, penulis menyadari bahwa penulisan ini masih sangat jauh dari kesempurnaan. Oleh karena itu, kritikan dan saran sangat Penulis harapkan. Semoga hal ini dapat bermanfaat bagi seluruh rekan-rekan pembaca. Amin.

Sumbawa, Desember 2021

Penyusun

ABSTRAK

Diabetes merupakan permasalahan yang serius yang dapat menyebabkan penyakit komplikasi hingga berakhir pada kematian. *World Health Organization* (WHO) diabetes membunuh 1,5 juta jiwa di tahun 2012. Dari permasalahan tersebut banyak penelitian di bidang *data mining* yang menerapkan metode klasifikasi menggunakan untuk melakukan diagnosis atau identifikasi terhadap penyakit diabetes. Penelitian ini bertujuan untuk mengetahui penerapan dan tingkat akurasi yang diperoleh algoritma KNN dalam klasifikasi penyakit diabetes serta mengetahui tingkat akurasi algoritma KNN setelah diterapkan BPSO sebagai seleksi fitur.

Penelitian ini dilakukan dengan tahapan-tahapan yaitu pengumpulan *dataset pima indian diabetes* sebagai *dataset* pengujian, *pre-processing data* menggunakan *min-max normalization*, pembagian *dataset* menggunakan metode *10-fold cross validation*, klasifikasi menggunakan algoritma KNN hingga proses pengujian untuk mengetahui tingkat akurasi yang diperoleh.

Hasil penelitian menunjukkan bahwa klasifikasi menggunakan algoritma KNN memperoleh akurasi terbaik yaitu sebesar 77,214% dengan diperoleh fitur yang berpengaruh pada *dataset pima* yaitu *Glucose*, *Blood Pressure*, *Skin Thickness*, *Insulin*, *BMI*, *Diabetes Pedigree Function*, dan. Dengan demikian algoritma KNN dapat digunakan untuk melakukan identifikasi penyakit diabetes karena memiliki tingkat akurasi yang jauh lebih baik.

Kata kunci: Klasifikasi Penyakit Diabetes, *K-Nearest Neighbor*

DAFTAR ISI

| | |
|---|----------|
| KATA PENGANTAR..... | i |
| BAB I PENDAHULUAN..... | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Batasan Masalah..... | 3 |
| 1.4 Tujuan dan Manfaat Penulisan | 4 |
| 1.4.1 Tujuan Penulisan..... | 4 |
| 1.4.2 Manfaat Penulisan..... | 4 |
| 1.5 Metodologi Penelitian | 5 |
| 1.5.1 Metode Pengumpulan Data | 5 |
| 1.5.2 Metode Pengolahan Data | 5 |
| 1.5.3 Metode yang digunakan | 5 |
| BAB II LANDASAN TEORI | 7 |
| 2.1 Penelitian Terkait..... | 7 |
| 2.2 Penyakit Diabetes | 11 |
| 2.3 <i>Data mining</i> | 11 |
| 2.3.1 Tahapan KDD | 12 |
| 2.3.2 Teknik <i>Data mining</i> | 14 |
| 2.4 K-Nearest Neighbor (KNN) | 16 |
| 2.5 Pre-Processing Data..... | 17 |
| 2.5.1 Normalisasi Data..... | 17 |
| 2.6 Uji Akurasi | 18 |
| 2.7 Python | 20 |
| 2.7.1 Jupyter Notebook | 20 |

| | | |
|--|--|-----------|
| 2.8 | RapidMiner | 21 |
| BAB III METODOLOGI PENELITIAN | | 22 |
| 3.1 | Tahapan Penelitian | 22 |
| 3.2 | Pengumpulan Data..... | 22 |
| 3.3 | Tahapan <i>Pre-Processing</i> Data..... | 23 |
| 3.3.1 | Tahapan Normalisasi..... | 23 |
| 3.4 | Pengujian Nilai K Optimal | 24 |
| 3.5 | Klasifikasi KNN | 24 |
| 3.5.1 | Contoh Proses Klasifikasi KNN Manual | 25 |
| 3.6 | Pengujian Kinerja | 29 |
| 3.6.1 | Tujuan Pengujian | 29 |
| 3.6.2 | Metode Pengujian..... | 30 |
| BAB IV HASIL DAN PEMBAHASAN..... | | 51 |
| 4.1 | Pengumpulan dataset | 51 |
| 4.2 | <i>Normalization</i> | 52 |
| BAB V PENUTUP | | 75 |
| 5.1 | Kesimpulan..... | 75 |
| 5.2 | Saran | 76 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2.1. Tahap – Tahap KDD (Mediana Aryuni, 2016) | 12 |
| Gambar 2.2 Ilustrasi algoritma KNN | 17 |
| Gambar 3.1 Proses Klasifikasi KNN | 23 |
| Gambar 4.1 Tahapan Pertama Proses Import Data | 52 |
| Gambar 4.2 Tahapan Kedua Proses Import Data | 53 |
| Gambar 4.3 Tahapan Ketiga Proses Import Data..... | 53 |
| Gambar 4.4 Tahapan Keempat Proses Import Data..... | 53 |
| Gambar 4.5 Tahapan Terakhir Proses Import Data | 54 |
| Gambar 4.6 Hasil Import Data | 54 |
| Gambar 4.7 Proses Normalisasi dengan Fitur Normalize pada RapidMiner | 55 |
| Gambar 4.9 Grafik Pengujian | 56 |

DAFTAR TABEL

| | |
|---|----|
| Tabel 2.1. Penelitian terkait | 7 |
| Tabel 2.2 Confussion matrix | 19 |
| Tabel 3.1 Sampel Data Training | 24 |
| Tabel 3.2 Sampel Data <i>Testing</i> | 25 |
| Tabel 3.3 Hasil Perangkingan Nilai <i>Euclidean</i> | 27 |
| Tabel 3.4 Nilai <i>Euclidean</i> Terbaik | 28 |
| Tabel 3.5 Hasil Klasifikasi Data <i>Testing</i> | 28 |
| Tabel 3.6 <i>Confussion Matrix</i> | 29 |
| Tabel 4.1 <i>Dataset pima indian diabetes</i> | 51 |

DAFTAR LAMPIRAN

| | |
|-------------------------------|-----|
| LAMPIRAN SAMPLE DATASET | A.1 |
|-------------------------------|-----|

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes merupakan masalah yang menyebabkan kadar gula darah dalam tubuh manusia meningkat pada tingkat yang lebih tinggi secara berkelanjutan (Choubey et al., 2020). Selain itu diabetes juga menjadi penyebab terjadinya penyakit komplikasi, seperti penyakit stroke, ginjal, jantung koroner, gangguan pada mata dan saraf, bahkan hingga mengakibatkan amputasi dan kematian (Maulida, 2020).

Menurut organisasi kesehatan dunia atau yang dikenal sebagai *World Health Organization* (WHO) diabetes membunuh 1,5 juta jiwa di tahun 2012. WHO juga memprediksi bahwa di tahun 2030 penyakit diabetes akan menjadi salah satu dari 7 faktor penyebab utama terjadinya kematian di dunia (KEMENKES RI, 2019). Maka untuk mengurangi angka kematian tersebut, yaitu dengan melakukan deteksi secara dini terhadap penderita diabetes untuk mengantisipasi terjadinya komplikasi yang disebabkan diabetes yang berujung pada kematian.

Menurut (Hairani et al., 2018) *Data mining* atau penambangan data menjadi salah satu metode yang bisa digunakan untuk melakukan prediksi atau identifikasi masalah. Dalam *data mining* terdapat beberapa metode yang dapat dimanfaatkan untuk melakukan prediksi atau identifikasi terhadap pasien penderita penyakit diabetes, salah satunya adalah metode klasifikasi.

Dalam laporan ini metode klasifikasi yang akan digunakan merupakan algoritma *lazy learning* yaitu algoritma *K-Nearest Neighbor* (KNN). Algoritma KNN merupakan salah satu algoritma untuk melakukan klasifikasi dari suatu kumpulan data. Dimana algoritma KNN melakukan klasifikasi objek berdasarkan nilai k atau tetangga terdekat (Saxena et al., 2014). Dalam laporan ini algoritma KNN akan digunakan untuk melakukan klasifikasi pada *dataset pima indian diabetes* yang diperoleh, karena algoritma KNN memiliki

beberapa kelebihan yaitu memiliki ketangguhan terhadap *training* data yang memiliki banyak noise dan efisien bila data yang digunakan besar (Yunus, 2018). Selain memiliki kelebihan seperti yang sudah dijelaskan algoritma KNN juga memiliki kelemahan yaitu kurang optimal dalam menentukan nilai k yang merupakan jumlah tetangga terdekat dan harus menentukan atribut yang akan dipilih atau seleksi fitur guna mendapat hasil terbaik (Mahardika et al., 2018). Dari adanya permasalahan tersebut maka dalam laporan ini permasalahan yang akan diatasi oleh penulis adalah seleksi fitur, guna meningkatkan kinerja algoritma KNN dalam melakukan klasifikasi penyakit diabetes.

Beberapa penelitian dibidang *data mining* juga telah dilakukan untuk mengatasi permasalahan penyakit diabetes. penelitian yang dilakukan (Maulida, 2020), dalam penelitiannya maulida mengukur kinerja algoritma *k-Nearest Neighbor* (KNN) dalam menangani *dataset* penderita diabetes. Penelitian (Wijanarto & Puspitasari, 2019) yang membahas tentang optimasi pada algoritma pengklasifikasian biner dengan menggunakan tuning parameter pada dataset penyakit diabetes mellitus. Penelitian (Yunita, 2016) yang melakukan klasifikasi penyakit diabetes menggunakan *k-Nearest Neighbor* (KNN). Penelitian (Mirqotussa'adah et al., 2017) yang menerapkan *discretization* dan teknik *bagging* untuk meningkatkan algoritma C4.5 dalam diagnosis penyakit diabetes. Dari beberapa penelitian yang sudah dilakukan khususnya yang menerapkan algoritma KNN, masih memiliki kekurangan yaitu tidak menerapkan metode seleksi fitur untuk memilih atribut yang relevan yang dapat mengurangi waktu komputasi pada metode klasifikasi.

Sedangkan penelitian yang menerapkan seleksi fitur dilakukan oleh (Yunus, 2018) dalam penelitiannya yunus menggunakan algoritma KNN untuk prediksi penyakit ginjal kronik.

Dari uraian di atas, maka dalam laporan ini penulis bertujuan untuk meningkatkan kinerja algoritma KNN dalam klasifikasi penyakit diabetes.

Dalam laporan ini data yang penulis gunakan untuk melakukan pengujian dalam laporan ini merupakan data pasien diabetes yang diperoleh dari situs kaggle.com pada *UCI Machine Learning PIMA Indians Diabetes*.

1.2 Rumusan Masalah

Adapun rumusan masalah dalam laporan ini berdasarkan pemaparan pada latar belakang diatas adalah sebagai berikut:

- a. Bagaimana tingkat akurasi yang diperoleh algoritma KNN dalam klasifikasi penyakit diabetes?
- b. Fitur apa saja yang diperoleh pada *dataset pima indian diabetes* yang akan berpengaruh terhadap hasil klasifikasi algoritma KNN?
- c. Bagaimana tingkat akurasi yang diperoleh algoritma KNN dalam klasifikasi penyakit diabetes?

1.3 Batasan Masalah

- a. *Dataset* yang digunakan dalam laporan ini adalah *dataset* public yaitu *Pima Indian Diabetes* yang penulis peroleh dari situs kaggle.com *UCI Machine Learning PIMA Indians Diabetes* dengan 8 variabel *independen* dan 1 variabel *dependen*.
- b. Dalam laporan ini algoritma KNN digunakan untuk melakukan klasifikasi berdasarkan dataset yang didapat.
- c. Proses validasi akan dilakukan dengan menggunakan metode *10-fold cross validation*.

1.4 Tujuan dan Manfaat Penulisan

1.4.1 Tujuan Penulisan

- a. Untuk mengetahui tingkat akurasi yang diperoleh algoritma KNN dalam klasifikasi penyakit diabetes.
- b. Untuk mengetahui fitur-fitur yang mempengaruhi hasil klasifikasi algoritma KNN.
- c. Untuk mengetahui tingkat akurasi algoritma KNN.

1.4.2 Manfaat Penulisan

- a. Dapat menambah pemahaman penulis terkait implementasi *data mining* khususnya penerapan metode klasifikasi menggunakan algoritma KNN serta dapat memahami proses *pre-processing* data.
- d. Bagi pengembangan ilmu, hasil dari laporan ini bisa menjadi bahan rujukan untuk pengembangan berikutnya tentang penerapan *data mining* khususnya metode klasifikasi.

1.5 Metodologi Penelitian

1.5.1 Metode Pengumpulan Data

Dalam laporan ini penulis menggunakan data public atau data sekunder yaitu data dari *PIMA Indian Diabetes* yang diperoleh dari *UCI Repository*.

1.5.2 Metode Pengolahan Data

Metode pengolahan data atau Pre-Processing data perlu dilakukan untuk menghilangkan atribut yang kurang lengkap, data noise, serta menangani data yang tidak konsisten.

1.5.3 Metode yang digunakan

Metode yang digunakan adalah algoritma K-Nearest Neighbor (KNN) dalam klasifikasi penyakit diabetes.

BAB II

LANDASAN TEORI

Landasan teori merupakan serangkain teori yang relevan yang dijadikan sebagai landasan dalam penelitian. Sehingga dalam bab ini akan dibahas hal-hal atau teori-teori yang berkaitan dengan permasalahan atau ruang lingkup pada laporan ini.

2.1 Penelitian Terkait

Berikut adalah beberapa penelitian terdahulu yang membahas tentang diabetes maupun algoritma KNN.

Tabel 2.1. Penelitian terkait

| No | Nama Peneliti | Topik | Metode yang digunakan |
|----|----------------------------------|---|---|
| 1 | Andi Maulida Argina | Penerapan Metode Klasifikasi K-Nearest Neighbor pada <i>Dataset</i> Penderita Penyakit Diabetes | <i>K-Nearest Neighbor</i> (KNN) |
| 2 | Wijanarto & Puspitasari, 2019 | Optimasi Algoritma Klasifikasi Biner dengan Tuning Parameter pada Penyakit Diabetes Mellitus | <i>Logistic Regression, Linear Discriminant Analysis, K-Nearest Neighbor, Decision Tree, Naïve Bayes, dan Support Vector Machine.</i> |

| | | | |
|---|------------------------------|---|--|
| 3 | Yunita, 2016 | Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor | K-Nearest Neighbor |
| 4 | Haerani et al., 2018 | Komparasi Akurasi Metode Correlated Naive Bayes Classifier Dan Naive Bayes Classifier Untuk Diagnosis Penyakit Diabetes | Correlated Naive Bayes Classifier Dan Naive Bayes Classifier |
| 5 | Mirqotussa'adah et al., 2017 | Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa Diabetes | C.45 |
| 6 | Liklikwatil et al., 2018 | Optimasi K-Nearest Neighbor Dengan Particle Swarm Optimization Untuk Memprediksi Harga Komoditi Karet | K-Nearest Neighbor dan Particle Swarm Optimization |

| | | | |
|---|------------------|--|--|
| 7 | Hasanuddin, 2016 | Perbandingan Algoritma Knn Dan Knn-Pso Untuk Klasifikasi Tingkat Pengetahuan Ibu Dalam Pemberian Asi Eksklusif | K-Nearest Neighbor dan Particle Swarm Optimization |
|---|------------------|--|--|

Penelitian yang dilakukan oleh (Maulida, 2020) yang berjudul Penerapan Metode Klasifikasi K-Nearest Neighbor Pada Dataset Penderita Penyakit Diabetes. Penelitian ini membahas tentang pengukuran performa metode klasifikasi KNN dalam mengelola *dataset* penderita diabetes. Hasil perhitungan nilai *accuracy* yang diperoleh dalam penelitian ini yaitu 39% pada $k=3$.

Penelitian yang dilakukan (Wijanarto & Puspitasari, 2019) yang berjudul Optimasi Algoritma Klasifikasi Biner dengan Tuning Parameter pada Penyakit Diabetes Mellitus. Penelitian ini membahas tentang optimasi algoritma klasifikasi biner pada penyakit diabetes melitus mulai dari observasi, visualisasi, *statistic* deskriptif *dataset*, *pre-processing dataset*, penentuan baseline model, tuning parameter model dan finalisasi model. Dimana algoritma yang digunakan adalah 3 algoritma linier yaitu *Logistic Regression*, *Linear Discriminant Analysis*, *K-Nearest Neighbor* dan 3 algoritma non-linier yaitu *Decision Tree*, *Naïve Bayes*, *Support Vector Machine*.

Penelitian yang dilakukan oleh (Yunita, 2016) yang berjudul Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor. Penelitian ini membahas tentang implementasi algoritma KNN untuk klasifikasi data penyakit diabetes mellitus. Hasil perhitungan nilai *accuracy* yang diperoleh dalam penelitian ini yaitu 96%.

Penelitian yang dilakukan oleh (Haerani et al., 2018) yang berjudul Komparasi Akurasi Metode Correlated Naïve Bayes Classifier Dan Naïve Bayes Classifier Untuk Diagnosis Penyakit Diabetes. Penelitian ini membahas tentang komparasi metode Correlated-Naïve Bayes Classifier dan Naïve Bayes Classifier untuk mendapatkan akurasi terbaik dalam 10uclidean penyakit diabetes. Dalam penelitian ini diperoleh hasil akurasi metode NBC sebesar 64,33% dan metode CNBC sebesar 67,15%.

Penelitian yang dilakukan oleh (Mirqotussa'adah et al., 2017) yang berjudul Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa Diabetes. Penelitian ini membahas tentang penerapan metode 10uclidean1010ion untuk menangani data 10uclidean10 dan 10uclid bagging pada algoritma C.45. Hasil nilai *accuracy* yang diperoleh dalam penelitian ini adalah 74,87%.

Penelitian yang dilakukan oleh (Liklikwatil et al., 2018) yang berjudul Optimasi K-Nearest Neighbor Dengan Particle Swarm Optimization Untuk Memprediksi Harga Komoditi Karet. Penelitian ini membahas tentang penerapan algoritma PSO pada algoritma KNN untuk prediksi time series harga komoditi karet spesifikasi teknis. Hasil yang diperoleh dalam penelitian ini, dengan menerapkan algoritma KNN, memperoleh nilai RMSE sebesar 0,082.

Penelitian yang dilakukan oleh (Hasanuddin, 2016) yang berjudul Perbandingan Algoritma Knn Dan Knn-Pso Untuk Klasifikasi Tingkat Pengetahuan Ibu Dalam Pemberian Asi Eksklusif. Penelitian ini membahas tentang prediksi tingkat pengetahuan ibu terhadap pemberian asi eksklusif dengan menggunakan algoritma KNN berbasis PSO. Hasil *accuracy* K-NN berbasis PSO yang diperoleh dalam penelitian ini adalah 74,36%.

Dari beberapa penelitian diatas terkait penyakit diabetes, belum ada penelitian yang menerapkan algoritma PSO untuk optimasi metode

klasifikasi yaitu algoritma KNN dalam menangani penyakit diabetes. Sehingga dalam laporan ini penulis akan menerapkan algoritma PSO untuk meningkatkan kinerja algoritma KNN dalam melakukan klasifikasi pada dataset *Pima Indians Diabetes*.

2.2 Penyakit Diabetes

Diabetes atau dikenal dengan penyakit kencing manis merupakan penyakit kronis serius yang disebabkan oleh meningkatnya kadar glukosa atau kadar gula diatas normal. Penyakit diabetes juga menjadi salah satu penyakit yang banyak diderita oleh manusia (Wijanarto & Puspitasari, 2019).

Dibetes melitus merupakan sesuatu yang tidak dapat dituangkan dalam satu jawaban yang jelas dan singkat, tapi secara umum dapat dikatakan sebagai suatu kumpulan problema 11uclidean dan kimiawi yang merupakan akibat dari sejumlah 11uclid (Dr. dr. Eva Decroli, 2019).

2.3 Data mining

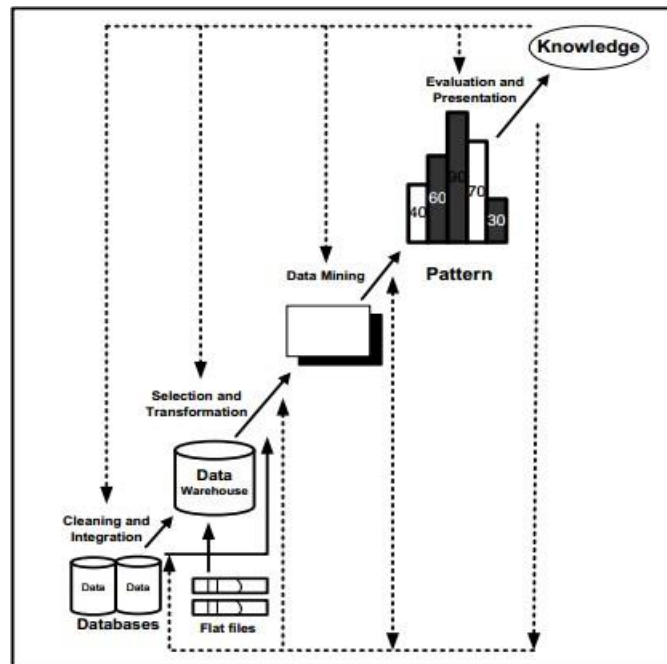
Data mining merupakan proses untuk menemukan korelasi baru yang bermakna, pola dan tren dengan memilah-milah sejumlah besar data yang tersimpan dalam repositori, menggunakan teknologi penalaran pola serta 11uclid-teknik 11uclidean dan matematika. Menurut (Hairani et al., 2018) *data mining*, juga merupakan suatu proses untuk menggali pengetahuan yang dibutuhkan dari sejumlah data besar (Yunus, 2018).

Data mining, sering juga disebut sebagai knowledge discovery in database (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santoso, 2007).

Dari dua pendapat diatas penulis dapat menyimpulkan bahwa *data mining* merupakan serangkaian proses yang menggunakan 11uclid matematis untuk menggali suatu informasi atau pengetahuan yang berguna untuk permasalahan di berbagai bidang dari suatu kumpulan data yang besar (*Big data*).

2.3.1 Tahapan KDD

Data mining dibagai menjadi beberapa tahapan yang di ilustrasikan pada Gambar 2.1.



Gambar 2.1. Tahap – Tahap KDD (Mediana Aryuni, 2016)

Tahap-tahap tersebut bersifat interaktif, pemakai terlibat langsung atau dengan perantaraan knowledge base Tahap-tahap *data mining* ada 6 yaitu (Mediana Aryuni, 2016):

1. Pembersihan data (*data cleaning*)

Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa *data mining* yang kita miliki. Data-data yang tidak relevan itu juga lebih baik dibuang karena keberadaannya bisa mengurangi mutu atau akurasi dari hasil *data mining* nantinya.

2. Integrasi data (*data integration*)

Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya

3. Transformasi data

Beberapa teknik *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa teknik standar seperti analisis asosiasi dan klustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut *binning*. Disini juga dilakukan pemilihan data yang diperlukan oleh teknik *data mining* yang dipakai. Transformasi dan pemilihan data ini juga menentukan kualitas dari hasil *data mining* nantinya karena ada beberapa karakteristik dari teknik-teknik *data mining* tertentu yang tergantung pada tahapan.

4. Teknik *data mining*

Teknik *data mining* sendiri hanya merupakan salah satu bagian dari proses KDD. Ada beberapa teknik *data mining* yang sudah umum dipakai. Kita akan membahas lebih jauh mengenai teknik-teknik yang ada di seksi berikutnya. Perlu diperhatikan bahwa ada kalanya teknik-teknik *data mining* umum yang tersedia di pasar tidak mencukupi untuk melaksanakan *data mining* di bidang tertentu atau untuk data tertentu. Sebagai contoh akhir-akhir ini dikembangkan berbagai teknik *data mining* baru untuk penerapan di bidang bioinformatika seperti analisa hasil microarray untuk mengidentifikasi DNA dan fungsi-fungsinya.

5. Evaluasi pola 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. Bila ternyata hasil yang diperoleh

tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti: menjadikannya umpan balik untuk memperbaiki proses *data mining*, mencoba teknik *data mining* lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

6. Interpretasi hasil

Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisa 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*.

2.3.2 Teknik *Data mining*

Menurut Pramudiono, 2003 ada tiga Teknik dalam *data mining* yang 14uclide yaitu:

1. *Association Rule Mining*

Association Rule Mining merupakan suatu teknik yang digunakan untuk menemukan aturan assosiasi antara suatu kombinasi item.

2. *Clustering*

Clustering merupakan pengelompokan data tanpa berdasarkan kelas data tertentu. Prinsip dari clustering adalah memaksimalkan kesamaan antar anggota dalam satu kelas dan meminimumkan kesamaan antar kelas.

3. Klasifikasi (*Classification*)

Klasifikasi merupakan suatu proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam melakukan klasifikasi dataset yang digunakan harus memiliki label atau atribut tujuan. Meramal objek kelas

pada setiap persoalan dalam data merupakan tujuan dari klasifikasi (Putra and Chan, 2018).

Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning (fase training)*, dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data tes digunakan untuk memperkirakan akurasi dari *rule* klasifikasi (dalam Rohman, 2012).

Algoritma klasifikasi menggunakan *data training* untuk membuat sebuah model. Model yang sudah dibangun tersebut kemudian digunakan untuk memprediksi label kelas data baru yang belum diketahui (Yunita, 2016). Selain itu dalam klasifikasi terdapat metode-metode atau model-model yang telah dikembangkan oleh peneliti untuk menyelesaikan kasus klasifikasi antara lain (Mardi & Yulia, 2014):

- a. Pohon keputusan
- b. Pengklasifikasi bayes / naive bayes
- c. Jaringan saraf tiruan
- d. Analisis statistik
- e. Algoritma genetika
- f. Rough sets
- g. Pengklasifikasi *K-Nearest Neighbor* (KNN)
- h. Metode berbasis aturan
- i. Memory based reasoning

Setiap algoritma dalam klasifikasi *data mining* tersebut memiliki kelebihan dan kekurangan. Tetapi prinsip dari masing-masing algoritma tersebut sama, yaitu melakukan suatu pelatihan sehingga di akhir pelatihan model dapat memprediksi setiap 15uclid masukan ke label kelas *output* dengan tepat (Yunita, 2016).

Dari penjelasan 16uclid-teknik *data mining* diatas, yang digunakan sebagai pengujian dalam laporan ini adalah 16uclid klasifikasidengan metode atau algoritma *K-Nearest Neighbor* (KNN).

2.4 K-Nearest Neighbor (KNN)

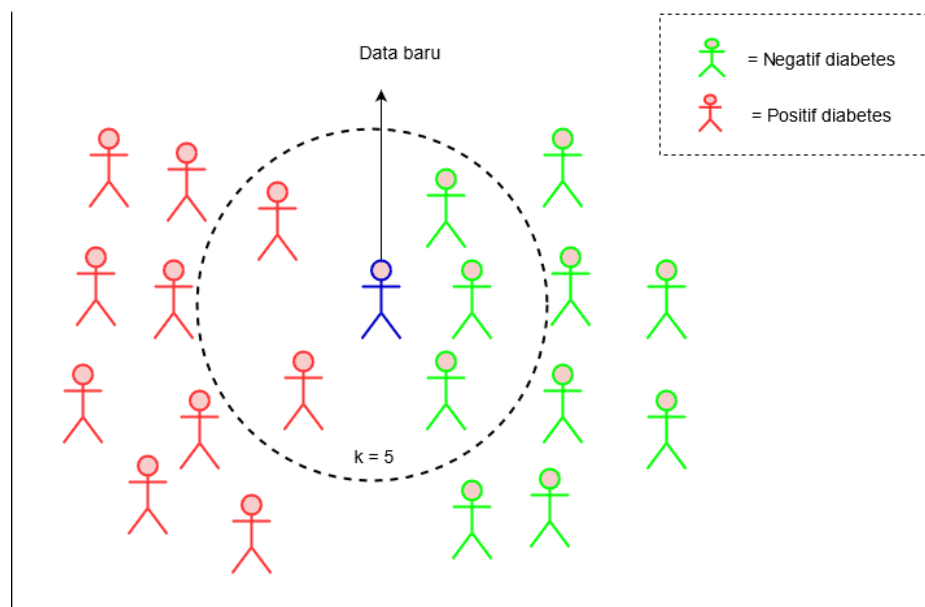
K-Nearest Neighbor (KNN) termasuk dalam kelompok *instance-based learning* dan merupakan salah satu *lazy learning* (Yunus, 2018). KNN juga merupakan salah satu metode klasifikasi dengan mengklasifikasikan objek yang jaraknya paling dekat. Prinsip kerja algoritma KNN adalah mencari jarak terdekat antara data yang akan di evaluasi dengan k tetangga (*neighbor*) dalam data pelatihan (Whidhiasih et al., 2013). *K-Nearest Neighbor* dapat memberikan keputusan untuk mengklasifikasikan data dari data latih dan mendapatkan hasil yang baik jika menggunakan data dalam jumlah yang besar (Saxena et al., 2014).

K-Nearest Neighbor memiliki kelebihan yaitu dapat menghasilkan data yang kuat atau jelas dan efektif jika digunakan pada data yang besar. Selain itu KNN juga memiliki kekurangan yaitu membutuhkan nilai k, jarak dari data percobaan tidak jelas dengan tipe jarak yang digunakan, untuk memperoleh hasil yang terbaik, maka harus menggunakan semua atribut atau hanya satu atribut yang telah pasti. Pemilihan nilai K (jumlah data/tetangga terdekat) ditentukan oleh peneliti. Pemilihan nilai K ini bisa mempengaruhi tingkat akurasi prediksi yang dikerjakan (Yunita, 2016).

Secara umum langkah-langkah algoritma KNN untuk mengklasifikasikan suatu data yaitu (Fernanda et al., 2017):

- a. Mendefinisikan nilai k.
- b. Melakukan perhitungan nilai jarak antara data latih dengan data uji.
- c. Mengelompokan data berdasarkan perhitungan jarak.
- d. Mengelompokan data berdasarkan nilai tetangga terdekat.
- e. Memilih nilai yang sering muncul dari tetangga terdekat sebagai hasil prediksi.

Berdasarkan uraian diatas maka algoritma KNN dapat diilustrasikan sebagai berikut:



Gambar 2.2 Ilustrasi algoritma KNN

2.5 Pre-Processing Data

Dalam *data mining* tahapan processing data merupakan tahapan yang perlu dilakukan sebelum melakukan tahapan pemodelan guna mendapatkan data yang berkualitas. Berikut adalah tahapan processing data yang akan dilakukan dalam laporan ini.

2.5.1 Normalisasi Data

Normalisasi data merupakan tahapan dalam metode transformasi data untuk melakukan penskalaan nilai atribut dari sekumpulan data sehingga memiliki rentang nilai tertentu. Dalam normalisasi data terdapat beberapa metode normalisasi yang dapat digunakan diantaranya adalah:

a. *Min-max normalization*

Min-max normalization merupakan metode normalisasi dengan melakukan transformasi linier terhadap data asli sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses.

b. *Z-score Normalization*

Z-score normalization merupakan metode normalisasi berdasarkan mean (nilai rata-rata) dan standar deviation (deviasi

standar) dari data. Metode ini akan sangat berguna jika nilai actual minimum dan maksimum dari data tidak diketahui.

c. Decimal Scaling Normalization

Decimal scaling merupakan metode normalisasi dengan menggerakkan nilai decimal data ke arah yang diinginkan.

2.6 Uji Akurasi

Akurasi klasifikasi merupakan ukuran ketepatan klasifikasi yang menunjukkan performansi teknik klasifikasi secara keseluruhan (Nugroho, dkk, 2003). Semakin tinggi akurasi klasifikasi berarti performansi teknik klasifikasi juga semakin baik. Kinerja metode klasifikasi dalam melakukan pengklasifikasian menggambarkan seberapa baik metode tersebut dalam mengklasifikasikan data.

Evaluasi model klasifikasi didasarkan pada pengujian untuk memperkirakan obyek yang benar dan salah, urutan pengujian ditabulasikan dalam *Confusion matrix* dimana kelas yang diprediksi ditampilkan dibagian atas matriks dan kelas yang diamati disisi kiri. Setiap sel berisi angka yang menunjukkan berapa banyak kasus yang sebenarnya dari kelas yang di amati untuk di prediksi (Yunus, 2018).

Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode kalsifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil kalsifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya (Prasetyo 2012). *Confussion matrix* berisi tentang perhitungan jumlah objek data testing yang diprediksikan kedalam sebuah kelas dengan klasifikasi yang sebenarnya (Kurniawan & Ivandari, 2017).

Secara umum bentuk *Confussion matrix* dapat dilihat pada Tabel 2.2 berikut:

Tabel 2.2 Confussion matrix

| <i>classification</i> | <i>Predicted Class</i> | | |
|-----------------------|------------------------|--------------------|-------------------|
| | | <i>Class = Yes</i> | <i>Class = No</i> |
| <i>Obevered Class</i> | <i>Class = Yes</i> | <i>TP</i> | <i>FN</i> |
| | <i>Class = No</i> | <i>FP</i> | <i>TN</i> |

Dalam confusion matrix terdapat total record yang dipakai dalam *dataset* baik yang diprediksikan kedalam kelas positif ataupun negatif. Tupel atau record dengan prediksi klasifikasi positif, prediksi klasifikasi negatif, serta kesalahan dalam klasifikasi dapat terlihat dalam matrix ini (Kurniawan & Ivandari, 2017). Untuk menghitung nilai accuracy, precision, dan recall dengan rumus perhitungan sebagai berikut:

$$Accurasi = \left(\frac{TP + TN}{TP + TN + FP + FN} \right) \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

Keterangan:

- a. TP merupakan *True Positive*, yaitu jumlah keseluruhan data positif yang terklasifikasi dengan benar.
- b. TN adalah True Negative, yaitu jumlah keseluruhan data negatif yang terklasifikasi dengan benar.
- c. FN merupakan False Negative, yaitu jumlah data negatif namun terklasifikasi salah.
- d. FP merupakan False Positive, yaitu jumlah data positif namun terklasifikasi salah.

2.7 Python

Menurut Hudaya (2013), Python adalah salah satu bahasa pemrograman *Open Source* yang ringkas, sederhana, dan bisa digunakan di beberapa sistem operasi. Python bersifat gratis dan bisa diunduh siapa saja melalui www.python.com dengan berbagai versi yang ada. Python sebenarnya sangat cocok untuk dipelajari oleh programmer pemula karena sifatnya yang mudah dan ringkas. Dikatakan mudah karena tidak perlu mendeklarasikan variabel seperti pemrograman Java, C, Pascal, dan sebagainya. Ia juga sangat ringkas. Untuk pemecahan masalah dengan Python dibutuhkan jumlah baris kode yang lebih sedikit dibandingkan dengan bahasa pemrograman lain. Salah satu ciri khas pemrograman Python adalah indentasi. Secara sederhana, indentasi adalah blok program yang ditandai dengan kesamaan jarak permulaan huruf kode dari ujung kiri layar. Secara asal, kode pada python harus dimulai penulisan hurufnya dari ujung kiri layer (Hudaya, 2013).

2.7.1 Jupyter Notebook

Jupyter membutuhkan Python untuk diinstal karena didasarkan pada bahasa Python. Terdapat beberapa alat yang akan mengotomatisasi instalasi Jupyter dari GUI, salah satunya software Anaconda. Jupyter Notebook pada dasarnya adalah file JSON dengan sejumlah anotasi. Bagian utama Jupyter Notebook yaitu metadata untuk mengatur dan menampilkan Notebook, nomor versi software yang digunakan untuk membuat Notebook, dan daftar sel (Toomey, 2016).

2.8 RapidMiner

RapidMiner adalah sebuah aplikasi atau *Software* perangkat lunak yang berfungsi sebagai alat pembelajaran pada ilmu data mining. *platform* dikembangkan oleh suatu perusahaan yang bertujuan untuk bisnis komersial, penelitian, pendidikan, pelatihan, serta semua langkah dalam pembelajaran yang menyangkut pada suatu data yang besar.

RapidMiner telah membuktikan telah sukses mencapai dan memberikan dampak bisnis yang cepat bagi lebih dari 40.000 organisasi di setiap industri untuk mendorong pendapatan, dan menghindari resiko.

1. Transparansi Penuh & Tata Kelola untuk Pembelajaran Mesin yang satu diantaranya memiliki peran *Easy to Tune* yang berarti Buka, periksa dan konfigurasi Ulang dari 1.500 + blok bangunan visual algoritma *Machine Learning* dan operator ilmu data.
2. Mudah Dijelaskan Platform otomatis membangun alur kerja analitik visual yang dikisahkan untuk berkolaborasi dengan pemangku kepentingan bisnis dan menawarkan metode yang kaya untuk penjelasan model
3. *End-to-end* merupakan hal yang mengenai tentang mengubah data dari sumber apapun, mengoptimasikan pemilihan dan validasi model *Machine Learning* terbaik.

BAB III

METODOLOGI PENELITIAN

Dalam bab ini akan dijelaskan tahapan-tahapan penelitian yang akan dilakukan dalam laporan ini. Dimulai dari pengumpulan data yang digunakan, hingga pada tahapan metode validasi yang digunakan untuk melakukan pengujian algoritma.

3.1 Tahapan Penelitian

Tahapan penelitian merupakan langkah-langkah atau gambaran yang akan digunakan penulis dalam laporan ini. tahapan penelitian ini dibuat dengan tujuan untuk memberikan kemudahan dalam melakukan penelitian.

3.2 Pengumpulan Data

Pada tahapan ini, dilakukan pengumpulan *dataset* yang sesuai dengan permasalahan pada laporan ini. *Dataset* yang digunakan dalam laporan ini merupakan data *public* yang diperoleh dari situs kaggle.com pada *UCI Machine Learning Pima Indians Diabetes. Dataset Pima Indians Diabetes* merupakan *dataset* yang berasal dari *National Institute Of Diabetes and Digestive and Kidney Diseasses*, dimana data ini berisi tentang 768 wanita dari suatu populasi. Dari hasil analisis terdapat 258 dinyatakan positif diabetes dan 500 dinyatakan negative. Selain itu dataset ini juga sangat umum digunakan oleh peneliti dibidang *data mining*. *Dataset* ini terdiri dari 768 *record* atau contoh data pasien terkena diabetes dan tidak terkena diabetes. Selain itu *dataset* ini memiliki 8 variabel prediktor yang merupakan faktor resiko penyakit diabetes yaitu *pregnancies*, *glucose*, *blood pressure*, *skin thickness*, *insulin*, *body mass index (BMI)*, *diabetes pedigree function* dan *age*. *Dataset pima* juga memiliki 1 variabel hasil atau label yaitu *Outcome*.

Dataset pima yang diperoleh akan dibagi menjadi dua bagian, yaitu data *training* dan data *testing*. Data training adalah data yang sudah diberikan pelabelan atau sudah memiliki kelas sebelumnya dan digunakan sebagai data pembelajaran, sedangkan data *testing* adalah data yang akan digunakan untuk melakukan pengujian atau merupakan data baru yang tidak diketahui kelas atau lebelnya sehingga akan ditentukan label atau kelasnya dengan melakukan klasifikasi.

3.3 Tahapan *Pre-Processing* Data

Pre-processing merupakan tahapan awal yang akan dilakukan dalam laporan ini setelah *dataset* diperoleh, dengan tujuan untuk memperoleh *dataset* yang berkualitas dan siap untuk dilakukan klasifikasi menggunakan algoritma KNN. *Dataset* *pima* yang diperoleh sebenarnya dapat digunakan langsung untuk klasifikasi, namun karena *dataset pima* yang diperoleh dari situs kaggle.com memiliki rentang nilai yang berbeda pada setiap atribut atau variabel *independen*, maka tahapan *pre-processing* data perlu dilakukan untuk mengatasi permasalahan tersebut sehingga dapat meningkatkan kinerja metode klasifikasi yang digunakan. *Pre-processing* data yang akan digunakan dalam laporan ini adalah normalisasi data dengan harapan dapat meningkatkan kinerja metode klasifikasi yang digunakan.

3.3.1 Tahapan Normalisasi

Tahapan *pre-processing* data yang akan dilakukan pertama kali adalah normalisasi data. Normalisasi data merupakan proses penskalaan nilai dari sekumpulan data agar berada pada rentang nilai tertentu. Normalisasi data dapat dilakukan menggunakan beberapa metode, salah satunya *min-max normalization*. *Min-max normalization* merupakan metode normalisasi dengan melakukan transformasi linier pada data asli. Penggunaan metode *min-max normalization* untuk melakukan penskalaan data pada laporan ini didasarkan pada penelitian yang dilakukan oleh (Nasution et al., 2019) dengan judul perbandingan normalisasi data untuk klasifikasi wine menggunakan algoritma KNN, dimana dalam penelitian tersebut hasil pengujian klasifikasi menggunakan algoritma KNN dengan normalisasi data menggunakan metode *min-max normalization* memperoleh hasil akurasi yang lebih baik dibandingkan dengan hasil klasifikasi menggunakan metode normalisasi yang lain, oleh karena itu dalam laporan ini penulis menggunakan metode *min-max normalization* dengan harapan akan memperoleh hasil yang baik juga.

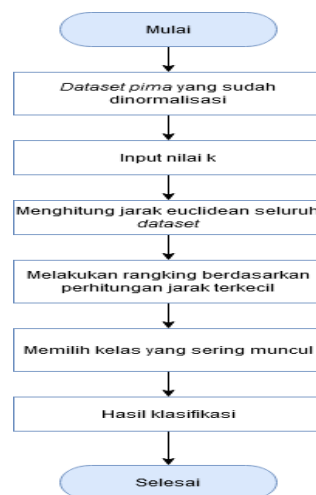
Dalam laporan ini keseluruhan atribut atau fitur pada *dataset pima* akan dilakukan normalisasi menggunakan metode *min-max normalization*, atribut tersebut terdiri dari *Pregnancies* (X1), *Glucose* (X2), *Blood Pressure* (X3), *Skin Thickness* (X4), *Insulin* (X5), *BMI* (X6), *Diabetes Pedigree Function* (X7), dan *Age* (X8).

3.4 Pengujian Nilai K Optimal

Dalam algoritma KNN penentuan nilai k sangat berpengaruh terhadap klasifikasi yang diperoleh, hal ini dikarenakan nilai k sangat sensitif terhadap hasil klasifikasi (Ulya et al., 2021). Dalam penelitian ini penetapan nilai k akan dilakukan dengan melakukan pengujian skor akurasi pada nilai k 5 sampai dengan nilai k 30. Dimana proses pencarian nilai k optimal akan dimulai dengan melakukan pembagian dataset menjadi data training dan data testing sebesar 70% dan 30% dari jumlah keseluruhan data, kemudian akan dilanjutkan dengan melakukan pemodelan menggunakan algoritma KNN dengan menetapkan nilai 5 sampai dengan 30 kemudian akan dilakukan pengujian skor akurasi yang diperoleh secara berulang sesuai dengan jumlah k yang ditetapkan. Dari perulangan akurasi tersebut maka akan diperoleh nilai atau skor akurasi pada masing-masing k yang digunakan. Nilai k yang memiliki skor akurasi tertinggi nantinya akan ditetapkan sebagai nilai k optimal yang akan digunakan untuk melakukan pengujian pada algoritma KNN. Hasil pengujian nilai akan dibahas pada bab selanjutnya.

3.5 Klasifikasi KNN

Pada tahap ini akan dilakukan proses klasifikasi menggunakan algoritma KNN dengan menggunakan nilai k optimal yang sudah diperoleh pada tahapan sebelumnya. Berikut adalah proses klasifikasi menggunakan algoritma KNN:



Gambar 3.1 Proses Klasifikasi KNN

Pada Gambar 3.1 menunjukkan bahwa tahapan awal klasifikasi menggunakan algoritma KNN dimulai dari penetapan *dataset* dimana *dataset* yang digunakan untuk melakukan proses klasifikasi menggunakan algoritma KNN merupakan *dataset pima* yang sudah melalui proses normalisasi data menggunakan *min-max normalization*, dimana *dataset* tersebut disimpan dalam bentuk csv. Pada tahapan kedua dalam proses klasifikasi menggunakan algoritma KNN akan dilakukan penentuan nilai k atau bisa juga disebut jumlah tetangga terdekat. Nilai k yang akan digunakan adalah nilai k optimal yang diperoleh pada tahapan sebelumnya. Nilai k yang menghasilkan kinerja terbaik nantinya akan digunakan untuk dijadikan sebagai pembanding antara algoritma KNN. Pada tahapan ketiga yaitu melakukan perhitungan jarak euclidean distance pada keseluruhan dataset dengan dataset baru yang akan dilakukan klasifikasi dengan menggunakan rumus pada persamaan 1 yang ada di bab sebelumnya. Kemudian tahapan selanjutnya yaitu melakukan perangkingan pada perhitungan euclidean yang didapatkan dan dilanjutkan dengan memilih kelas dengan jumlah kemunculan terbanyak berdasarkan jumlah k tetangga.

3.5.1 Contoh Proses Klasifikasi KNN Manual

Berikut adalah proses klasifikasi KNN secara manual dengan menghitung jarak ketetanggaan antara data *testing* dan data training. Proses klasifikasi secara manual akan dilakukan dengan menggunakan 10 data sebagai sampel untuk data *training* dan 1 data sebagai data *testing*. Berikut adalah data sampel yang digunakan:

Tabel 3.1 Sampel Data Training

| No | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 | X_7 | X_8 | Y |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| 1 | 0.353 | 0.744 | 0.590 | 0.354 | 0.000 | 0.501 | 0.234 | 0.483 | 1 |
| 2 | 0.059 | 0.427 | 0.541 | 0.293 | 0.000 | 0.396 | 0.117 | 0.167 | 0 |
| 3 | 0.471 | 0.920 | 0.525 | 0.000 | 0.000 | 0.347 | 0.254 | 0.183 | 1 |
| 4 | 0.059 | 0.447 | 0.541 | 0.232 | 0.111 | 0.419 | 0.038 | 0.000 | 0 |
| 5 | 0.000 | 0.688 | 0.328 | 0.354 | 0.199 | 0.642 | 0.944 | 0.200 | 1 |
| 6 | 0.294 | 0.583 | 0.607 | 0.000 | 0.000 | 0.382 | 0.053 | 0.150 | 0 |
| 7 | 0.176 | 0.392 | 0.410 | 0.323 | 0.104 | 0.462 | 0.073 | 0.083 | 1 |
| 8 | 0.588 | 0.578 | 0.000 | 0.000 | 0.000 | 0.526 | 0.024 | 0.133 | 0 |

| | | | | | | | | | |
|----|-------|-------|-------|-------|-------|-------|-------|-------|---|
| 9 | 0.118 | 0.990 | 0.574 | 0.455 | 0.642 | 0.455 | 0.034 | 0.533 | 1 |
| 10 | 0.471 | 0.628 | 0.787 | 0.000 | 0.000 | 0.000 | 0.066 | 0.550 | 1 |

Berdasarkan pada Tabel 3.1, dataset yang akan digunakan sebagai sampel data *training* merupakan 10 data awal pada dataset *pima* yang sebelumnya sudah dilakukan proses normalisasi menggunakan metode *min-max normalization* terlihat bahwa dari *dataset* tersebut sudah memiliki rentang nilai yang sama pada setiap atribut atau fitur.

Tabel 3.2 Sampel Data Testing

| No | X ₁ | X ₂ | X ₃ | X ₄ | X ₅ | X ₆ | X ₇ | X ₈ | Y |
|----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|---|
| 15 | 0.294 | 0.834 | 0.590 | 0.192 | 0.207 | 0.385 | 0.217 | 0.500 | ? |

Berdasarkan pada Table 3.1, dataset yang akan dijadikan sebagai data testing merupakan data ke 15 dari *dataset* *pima* yang sebelumnya sudah dilakukan normalisasi juga. Dari *dataset* tersebut akan dilakukan proses klasifikasi untuk mengetahui yang dijadikan sebagai data *testing* merupakan data pasien positif diabetes atau data pasien negatif diabetes menggunakan algoritma KNN dengan mencari jarak ketetanggaan menggunakan *euclidean distance*.

1. Nilai k=3 akan digunakan sebagai contoh untuk menentukan jumlah ketetanggaan terdekat.
2. Menghitung jarak ketetanggaan data *testing* dengan data training ke 1

$$d(1,15) =$$

$$\sqrt{(0.353 - 0.294)^2 + (0.744 - 0.834)^2 + (0.590 - 0.590)^2 + (0.354 - 0.192)^2 + (0.000 - 0.207)^2 + (0.501 - 0.385)^2 + (0.234 - 0.217)^2 + (0.483 - 0.500)^2}$$

$$= \sqrt{0.09464}$$

$$= 0.30763$$

3. Menghitung jarak ketetanggaan data *testing* dengan data training ke 2

$$d(2,15) =$$

$$\sqrt{(0.059 - 0.294)^2 + (0.427 - 0.834)^2 + (0.541 - 0.590)^2 + (0.293 - 0.192)^2 + (0.000 - 0.207)^2 + (0.501 - 0.385)^2 + (0.234 - 0.217)^2 + (0.167 - 0.500)^2}$$

4. Menghitung jarak ketetanggaan data *testing* dengan data training ke 3

$$d(3,15) =$$

$$\sqrt{(0.471 - 0.294)^2 + (0.920 - 0.834)^2 + (0.525 - 0.590)^2 + (0.000 - 0.192)^2 + (0.000 - 0.207)^2 + (0.347 - 0.385)^2 + (0.254 - 0.217)^2 + (0.183 - 0.500)^2}$$

$$= \sqrt{0.22535}$$

$$= 0.47471$$

5. Menghitung jarak ketetanggaan data *testing* dengan data training ke 4

$$d(4,15) =$$

$$\sqrt{(0.059 - 0.294)^2 + (0.447 - 0.834)^2 + (0.541 - 0.590)^2 + (0.232 - 0.192)^2 + (0.111 - 0.207)^2 + (0.419 - 0.385)^2 + (0.038 - 0.217)^2 + (0.000 - 0.500)^2}$$

$$= \sqrt{0.50164}$$

$$= 0.70826$$

6. Menghitung jarak ketetanggaan data *testing* dengan data training ke 5

$$d(5,15) =$$

$$\sqrt{(0.000 - 0.294)^2 + (0.688 - 0.834)^2 + (0.328 - 0.590)^2 + (0.354 - 0.192)^2 + (0.199 - 0.207)^2 + (0.642 - 0.385)^2 + (0.944 - 0.217)^2 + (0.200 - 0.500)^2}$$

$$= \sqrt{0.88671}$$

$$= 0.94165$$

7. Menghitung jarak ketetanggaan data *testing* dengan data training ke 6

$$d(6,15) =$$

$$\sqrt{(0.294 - 0.294)^2 + (0.583 - 0.834)^2 + (0.607 - 0.590)^2 + (0.000 - 0.192)^2 + (0.000 - 0.207)^2 + (0.382 - 0.385)^2 + (0.053 - 0.217)^2 + (0.150 - 0.500)^2}$$

8. Menghitung jarak ketetanggaan data *testing* dengan data training ke 7

$$d(7,15) =$$

$$\sqrt{(0.176 - 0.294)^2 + (0.392 - 0.834)^2 + (0.410 - 0.590)^2 + (0.323 - 0.192)^2 + (0.104 - 0.207)^2 + (0.462 - 0.385)^2 + (0.073 - 0.217)^2 + (0.083 - 0.500)^2}$$

$$= \sqrt{0.47030}$$

$$= 0.68578$$

9. Menghitung jarak ketetanggaan data *testing* dengan data training ke 8

$$d(8,15) =$$

$$\begin{aligned}
& (0.588 - 0.294)^2 + (0.578 - 0.834)^2 + (0.000 - 0.590)^2 + \\
& \sqrt{(0.000 - 0.192)^2 + (0.000 - 0.207)^2 + (0.526 - 0.385)^2 +} \\
& (0.024 - 0.217)^2 + (0.133 - 0.500)^2 \\
& = \sqrt{0.77201} \\
& = 0.87864
\end{aligned}$$

10. Menghitung jarak ketetangaan data *testing* dengan data training ke 9
 $d(9,15)$

$$\begin{aligned}
& (0.118 - 0.294)^2 + (0.990 - 0.834)^2 + (0.574 - 0.590)^2 + \\
& = \sqrt{(0.455 - 0.192)^2 + (0.642 - 0.207)^2 + (0.455 - 0.385)^2 +} \\
& (0.034 - 0.217)^2 + (0.533 - 0.500)^2 \\
& = \sqrt{0.35343} \\
& = 0.59450
\end{aligned}$$

11. Menghitung jarak ketetangaan data *testing* dengan data training ke 10
 $d(10,15)$

$$\begin{aligned}
& (0.471 - 0.294)^2 + (0.628 - 0.834)^2 + (0.787 - 0.590)^2 + \\
& = \sqrt{(0.000 - 0.192)^2 + (0.000 - 0.207)^2 + (0.000 - 0.385)^2 +} \\
& (0.066 - 0.217)^2 + (0.550 - 0.500)^2 \\
& = \sqrt{0.36523} \\
& = 0.60435
\end{aligned}$$

12. Melakukan perangkingan nilai *euclidean*

Setelah proses perhitungan *euclidean* dan jarak ketetangaan antara data *testing* dengan data training diperoleh, maka proses selanjutnya yaitu melakukan pengurutan nilai *euclidean* dengan cara melakukan perangkingan secara *ascending* dimana nilai *euclidean* terkecil akan diberikan nilai 1. Berikut adalah hasil perangkingan nilai *euclidean* disajikan dalam Tabel 3.2.

Tabel 3.3 Hasil Perangkingan Nilai Euclidean

| Data | Jarak Euclidean Distance | Ranking | Y |
|------|--------------------------|---------|---|
| 1 | 0.30763 | 1 | 1 |
| 2 | 0.63076 | 6 | 0 |
| 3 | 0.47471 | 2 | 1 |
| 4 | 0.70826 | 8 | 0 |
| 5 | 0.94165 | 10 | 1 |
| 6 | 0.54102 | 3 | 0 |

| | | | |
|----|---------|---|---|
| 7 | 0.68578 | 7 | 1 |
| 8 | 0.87864 | 9 | 0 |
| 9 | 0.59450 | 4 | 1 |
| 10 | 0.60435 | 5 | 1 |

13. Memilih kelas mayoritas

Dari Tabel 3.2 diperoleh 3 nilai *euclidean* terbaik berdasarkan nilai $k = 3$ yaitu 0.30763, 0.63076 dan 0.47471. Oleh karena itu maka selanjutnya akan dilakukan proses memilih kelas mayoritas untuk dijadikan sebagai hasil klasifikasi.

Tabel 3.4 Nilai Euclidean Terbaik

| Data | Jarak Euclidean Distance | Ranking | Y |
|------|--------------------------|---------|---|
| 1 | 0.30763 | 1 | 1 |
| 3 | 0.47471 | 2 | 1 |
| 6 | 0.54102 | 3 | 0 |

14. Hasil klasifikasi

Pada Tabel 3.4 diperoleh mayoritas kelas yang muncul yaitu kelas 1 (positive). Maka hasil klasifikasi terhadap data *testing* adalah kelas positive diabetes. Berikut adalah hasil klasifikasi *dataset pima indian diabetes* pada baris ke-15.

Tabel 3.5 Hasil Klasifikasi Data Testing

| No | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 | X_7 | X_8 | Y |
|----|-------|-------|-------|-------|-------|-------|-------|-------|---|
| 15 | 0.294 | 0.834 | 0.590 | 0.192 | 0.207 | 0.385 | 0.217 | 0.500 | 1 |

Berdasarkan Tabel 3.9 menunjukkan bahwa hasil klasifikasi dataset pima indian diabetes pada baris ke-15 adalah pasien positif diabetes.

3.6 Pengujian Kinerja

3.6.1 Tujuan Pengujian

Pada tahap ini akan dilakukan proses pengujian kinerja untuk mengetahui kinerja yang diperoleh pada algoritma KNN dalam melakukan klasifikasi penyakit diabetes berdasarkan pada dataset yang diperoleh. Pengujian dilakukan dengan mengevaluasi perbandingan hasil *Accuracy*,

diamana semakin tinggi nilai *Accuracy* yang diperoleh maka semakin baik pula metode atau algoritma yang digunakan.

3.6.2 Metode Pengujian

3.6.2.1 Confussion Matrix

Terdapat 4 istilah untuk mempresentasikan hasil proses klasifikasi. Dalam *confussion matrix*, terlihat pada Tabel 3.4.

Tabel 3.6 Confussion Matrix

| <i>classification</i> | <i>Predicted Class</i> | | |
|-----------------------|------------------------|-----------------|-----------------|
| | | <i>positive</i> | <i>negative</i> |
| <i>Obevered Class</i> | <i>positive</i> | <i>TP</i> | <i>FN</i> |
| | <i>negative</i> | <i>FP</i> | <i>TN</i> |

Keterangan:

TP: True Positive (Jumlah prediksi benar pada kelas positif) FP: False Positive (Jumlah prediksi salah pada kelas positif) FN: False Negative (Jumlah prediksi salah pada kelas negatif) TN: True Negative (Jumlah prediksi benar pada kelas negatif)

Berdasarkan dari 4 nilai klasifikasi tersebut, akan diperoleh nilai akurasi, presisi, dan recall. Nilai akurasi merupakan nilai yang digunakan untuk menggambarkan seberapa akurat metode klasifikasi yang digunakan dalam melakukan klasifikasi dengan benar.

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan mengenai hasil dari penelitian yang dilakukan serta pembahasan penelitian. Pembahasan penelitian meliputi hasil *pre-processing* data menggunakan *min-max normalization*, hasil klasifikasi algoritma KNN, hasil akurasi algoritma KNN yang.

4.1 Pengumpulan dataset

Seperti yang sudah dijelaskan pada bab sebelumnya dataset yang digunakan untuk penelitian dalam laporan ini merupakan dataset yang diperoleh dari situs kaggle.com yang berjumlah 768 data dengan 8 variabel *independen* dan 1 variabel *dependen*. Dimana data ini akan digunakan untuk melakukan pengujian menggunakan algoritma *K-Nearest*. Berikut adalah penampakan dataset pima yang akan digunakan untuk pengujian.

Tabel 4.1 Dataset pima indian diabetes

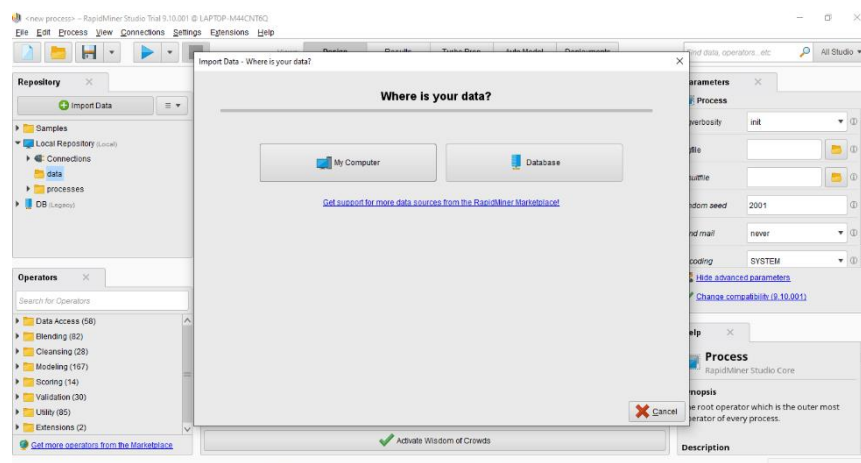
| No | X ₁ | X ₂ | X ₃ | X ₄ | X ₅ | X ₆ | X ₇ | X ₈ | Y |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|-----|
| 1 | 6 | 148 | 72 | 35 | 0 | 33.6 | 0.627 | 50 | 1 |
| 2 | 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 | 0 |
| 3 | 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 | 1 |
| 4 | 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 | 21 | 0 |
| 5 | 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 | 33 | 1 |
| 6 | 5 | 116 | 74 | 0 | 0 | 25.6 | 0.201 | 30 | 0 |
| 7 | 3 | 78 | 50 | 32 | 88 | 31 | 0.248 | 26 | 1 |
| 8 | 10 | 115 | 0 | 0 | 0 | 35.3 | 0.134 | 29 | 0 |
| 9 | 2 | 197 | 70 | 45 | 543 | 30.5 | 0.158 | 53 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 764 | 10 | 101 | 76 | 48 | 180 | 32.9 | 0.171 | 63 | 0 |
| 765 | 2 | 122 | 70 | 27 | 0 | 36.8 | 0.34 | 27 | 0 |

| No | X ₁ | X ₂ | X ₃ | X ₄ | X ₅ | X ₆ | X ₇ | X ₈ | Y |
|-----|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|---|
| 766 | 5 | 121 | 72 | 23 | 112 | 26.2 | 0.245 | 30 | 0 |
| 767 | 1 | 126 | 60 | 0 | 0 | 30.1 | 0.349 | 47 | 1 |
| 768 | 1 | 93 | 70 | 31 | 0 | 30.4 | 0.315 | 23 | 0 |

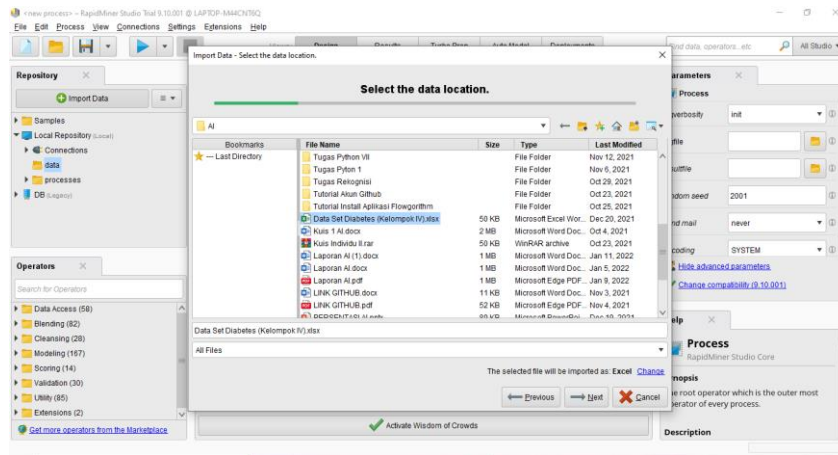
Dari Tabel 4.1 terlihat bahwa dataset pima yang diperoleh belum dilakukan tahapan normalisasi dimana pada kolom X7 pada data tersebut memiliki rentang nilai yang berbeda dengan data pada kolom lain, sehingga hal ini akan berpengaruh terhadap akurasi klasifikasi oleh algoritma KNN. Selain itu dataset tersebut masih memiliki fitur atau atribut yang lengkap yaitu berjumlah 8 atribut dan 1 label atau variabel dependen dimana hal ini akan berpengaruh terhadap waktu komputasi dan akurasi yang dihasilkan oleh algoritma KNN dikarenakan algoritma KNN akan melakukan training keseluruhan data yang ada yang berjumlah 768 data dan 8 atribut tersebut.

4.2 Normalization

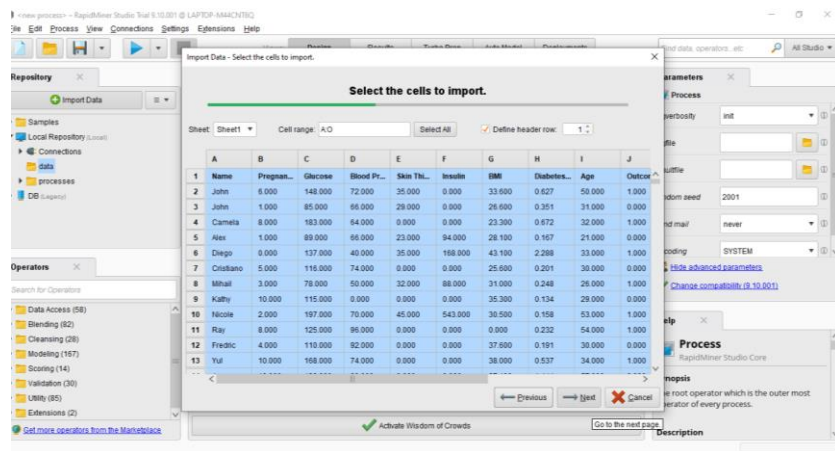
Dalam laporan ini tahapan pengujian baik *processing* data, klasifikasi hingga proses uji validasi atau uji akurasi akan dilakukan dengan memanfaatkan tools *rapid miner*, oleh karena itu untuk mempermudah penulis dalam melakukan pengujian pada laporan ini, penulis akan memanfaatkan rapid miner sebagai berikut:



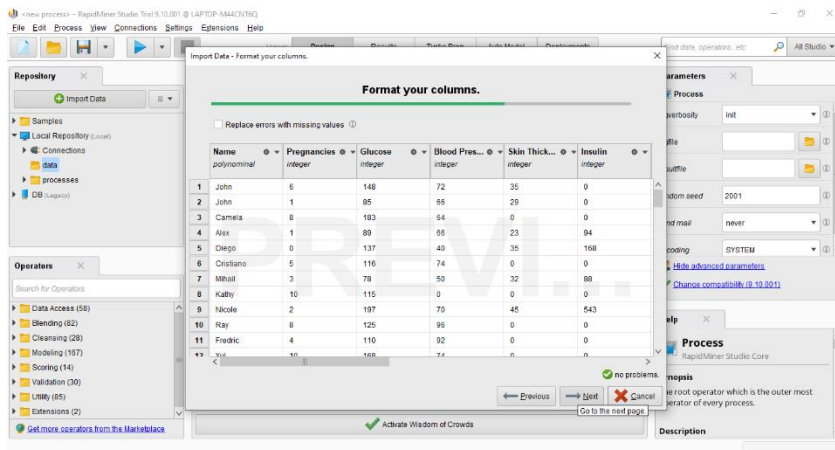
Gambar 4.1 Tahapan pertama proses import data



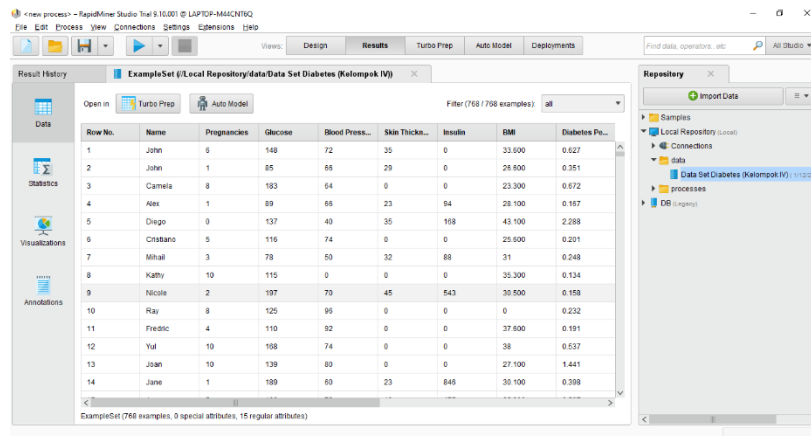
Gambar 4.2 Tahapan kedua proses import data



Gambar 4.3 Tahapan ke tiga proses import data



Gambar 4.4 Tahapan ke empat proses import data



Gambar 4.5 Tahapan terakhir proses import data

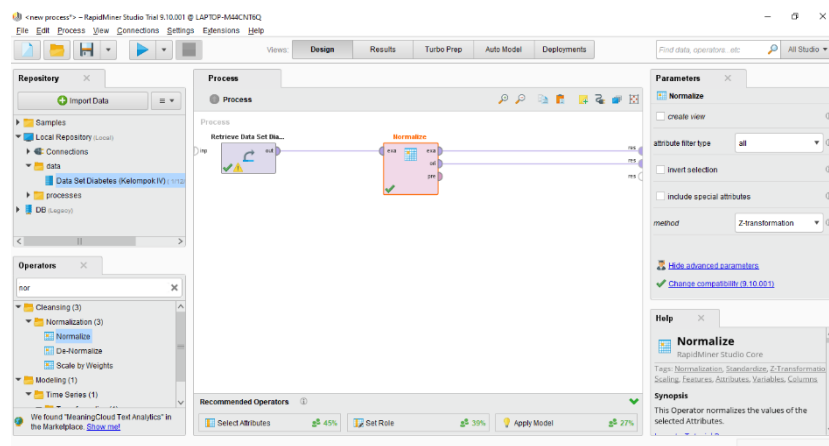
Dari Gambar 4.1 sampai 4.5 terlihat bahwa banyak fungsi dari *rapid miner* yang dapat digunakan. Berikut adalah hasil dari input data menggunakan *rapid miner*.

| Row No. | Name | Pregnancies | Glucose | Blood Press... | Skin Thicken... | Insulin | BMI | Diabetes Pe... |
|---------|-----------|-------------|---------|----------------|-----------------|---------|--------|----------------|
| 1 | John | 6 | 148 | 72 | 35 | 0 | 33.600 | 0.627 |
| 2 | John | 1 | 85 | 66 | 29 | 0 | 26.600 | 0.351 |
| 3 | Camela | 8 | 183 | 64 | 0 | 0 | 23.300 | 0.672 |
| 4 | Alex | 1 | 89 | 66 | 23 | 94 | 28.100 | 0.167 |
| 5 | Diego | 0 | 137 | 40 | 35 | 168 | 43.100 | 2.288 |
| 6 | Cristiano | 5 | 116 | 74 | 0 | 0 | 25.600 | 0.201 |
| 7 | Mihail | 3 | 78 | 50 | 32 | 88 | 31 | 0.248 |
| 8 | Kathy | 10 | 115 | 0 | 0 | 0 | 35.300 | 0.134 |
| 9 | Nicole | 2 | 197 | 70 | 45 | 543 | 30.500 | 0.158 |
| 10 | Ray | 8 | 125 | 96 | 0 | 0 | 0 | 0.232 |
| 11 | Fredric | 4 | 110 | 92 | 0 | 0 | 37.600 | 0.191 |
| 12 | Yul | 10 | 168 | 74 | 0 | 0 | 38 | 0.537 |
| 13 | Joan | 10 | 139 | 80 | 0 | 0 | 27.100 | 1.441 |
| 14 | Jane | 1 | 189 | 60 | 23 | 846 | 30.100 | 0.398 |

Gambar 4.6 Hasil Import Data

Berdasarkan Gambar 4.6 data yang diperoleh dari situs kaggle.com terdiri dari 8 variabel *independen*. Variabel *independen* pada data tersebut merupakan faktor terjadinya diabetes sedangkan variabel dependen merupakan variabel hasil atau prediksi dimana pasien tersebut positif diabetes atau tidak. Variabel *independen* pada data tersebut masih memiliki rentang nilai yang berbeda sehingga akan berpengaruh terhadap hasil klasifikasi yang diperoleh. Oleh karena itu penulis akan melakukan tahapan normalisasi menggunakan metode *min-max normalization* dengan memanfaatkan salah satu fungsi yang tersedia pada rapid miner yaitu *Normalize*.

Berikut adalah proses normalisasi menggunakan fitur *normalize* pada *RapidMiner*.



Gambar 4.7 Proses normalisasi dengan fitur *normalize* pada *rapidminer*

Berikut adalah hasil normalisasi dari proses yang ditunjukkan oleh Gambar 4.6, disajikan dalam Tabel 4.1 berikut:

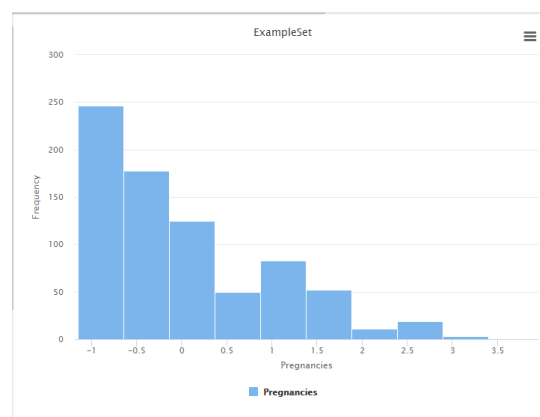
| Row No. | Pregnancies | Glucose | Blood Press... | Skin Thickn... | Insulin | BMI | Diabetes Pe... | Age |
|---------|-------------|---------|----------------|----------------|---------|--------|----------------|--------|
| 1 | 0.640 | 0.848 | 0.150 | 0.907 | -0.692 | 0.204 | 0.468 | 1.425 |
| 2 | -0.844 | -1.123 | -0.160 | 0.531 | -0.692 | -0.684 | -0.365 | -0.191 |
| 3 | 1.233 | 1.942 | -0.264 | -1.287 | -0.692 | -1.103 | 0.604 | -0.106 |
| 4 | -0.844 | -0.998 | -0.160 | 0.154 | 0.123 | -0.494 | -0.920 | -1.041 |
| 5 | -1.141 | 0.504 | -1.504 | 0.907 | 0.765 | 1.409 | 5.481 | -0.020 |
| 6 | 0.343 | -0.153 | 0.253 | -1.287 | -0.692 | -0.811 | -0.818 | -0.276 |
| 7 | -0.251 | -1.342 | -0.987 | 0.719 | 0.071 | -0.126 | -0.676 | -0.616 |
| 8 | 1.827 | -0.184 | -3.570 | -1.287 | -0.692 | 0.420 | -1.020 | -0.361 |
| 9 | -0.548 | 2.380 | 0.046 | 1.534 | 4.019 | -0.189 | -0.947 | 1.680 |
| 10 | 1.233 | 0.128 | 1.389 | -1.287 | -0.692 | -4.058 | -0.724 | 1.765 |
| 11 | 0.046 | -0.341 | 1.183 | -1.287 | -0.692 | 0.711 | -0.848 | -0.276 |
| 12 | 1.827 | 1.473 | 0.253 | -1.287 | -0.692 | 0.762 | 0.197 | 0.065 |
| 13 | 1.827 | 0.566 | 0.563 | -1.287 | -0.692 | -0.621 | 2.925 | 2.020 |

Gambar 4.8 Hasil normalisasi data dengan fitur *normalize* pada *rapidminer*

Berdasarkan Gambar 4.8 merupakan dataset *pima* setelah dilakukan normalisasi terlihat bahwa pada setiap kolom atau setiap variabel *independent* pada dataset *pima* setelah dilakukan tahapan normalisasi menggunakan *rapidminer* dengan memanfaatkan fungsi *Normalize* pada *rapidminer*, memiliki rentang nilai yang sama yaitu berada pada rentang nilai 0 sampai 1, dimana sebelumnya pada dataset *pima* memiliki rentang nilai yang berbeda pada setiap atribut atau variabel *independent*.

normalization untuk memperoleh data yang memiliki rentang nilai yang sama pada setiap data sudah tercapai dan kemungkinan akan meningkatkan akurasi yang diperoleh oleh algoritma KNN dalam melakukan klasifikasi pada *dataset* tersebut.

Dari tabel 4.1 diatas maka diperoleh nilai k optimal yaitu 19 dengan score akurasi sebesar 77,489%. Berikut adalah grafik perbandingan akurasi dalam pencarian nilai k optimal yang disajikan dalam Gambar 4.8.



Gambar 4.9 Grafik Pengujian

Berdasarkan Tabel 4.1 dan Gambar 4.6 memperlihatkan hasil pengujian dengan algoritma KNN dan menghasilkan akurasi tertinggi sebesar 77,489% pada $k=19$. Selain itu tampak pada grafik yang ditunjukkan oleh gambar 4.6 bahwa semakin tinggi nilai k yang digunakan hasil yang diperoleh atau akurasi yang diperoleh cenderung menurun, walaupun nilai k terendah didapatkan pada nilai $k=6$ dengan score akurasi sebesar 72,294%. Hal ini menunjukkan bahwa penggunaan nilai k yang digunakan mempengaruhi tingkat akurasi yang dihasilkan, hal ini juga selaras dengan beberapa penelitian yang mengatakan bahwa nilai k terbaik atau k optimal yaitu berada pada $k=3$ hingga $k=19$. Selain itu secara umum, penggunaan nilai k yang tinggi akan mengurangi efek noise pada klasifikasi, namun akan menyebabkan Batasan antar setiap klasifikasi menjadi lebih kabur. Berdasarkan hasil pengujian yang sudah dilakukan pada nilai $k=19$ yang merupakan nilai k optimal yang diperoleh dalam penelitian, nilai k tersebut akan penulis tetapkan untuk melakukan klasifikasi pada *dataset pima indian diabetes* yang sudah dilakukan menggunakan algoritma KNN dan rapidminer.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil dan pembahasan pada proses klasifikasi pada *dataset pima indian diabetes* menggunakan algoritma KNN diperoleh hasil klasifikasi 192 data terklasifikasi positif diabetes dan 576 data terklasifikasi negatif diabetes dan diperoleh tingkat akurasi atau tingkat korelasi antara nilai prediksi dari sistem dengan nilai sesungguhnya yang mencapai 75%.

Dari hasil pengujian menggunakan *confussion matrix* diperoleh peningkatan akurasi klasifikasi pada algoritma menjadi 77,214% sehingga tampak selisih akurasi yang diperoleh yaitu sebesar 2,214%. Algoritma KNN dapat digunakan untuk melakukan identifikasi penyakit diabetes.

5.2 Saran

Adapun saran-saran pengembangan lebih lanjut dari laporan ini yang dapat penulis berikan adalah sebagai berikut:

1. Pada penelitian lebih lanjut diharapkan selain menggunakan algoritma KNN kita dapat menggunakan algoritma – algoritma yang lain salah satunya PSO, hal ini disebabkan nilai k pada algoritma KNN sangat berpengaruh terhadap keakuratan klasifikasi algoritma KNN.
2. Diharapkan pada pengembangan lebih lanjut dapat menggunakan PSO untuk optimasi pada algoritma klasifikasi lain seperti SVM, *Naïve Bayes*, C.45, dan lain-lain guna mendapatkan algoritma yang lebih akurat dan efisien dalam melakukan klasifikasi penyakit diabetes. Selain itu juga dapat menggunakan beberapa algoritma untuk optimasi seperti GA,
3. Diharapkan dapat menggunakan metode yang lebih spesifik untuk mengukur dan memvalidasi kinerja algoritma yang dihasilkan, seperti menggunakan kurva AUC atau ROC.

DAFTAR REFERENSI

- Achyani, Y. E. (2018). Penerapan Metode Particle Swarm Optimization Pada Optimasi Prediksi Pemasaran Langsung. *Jurnal Informatika*, 5(1), 1–11. <https://doi.org/10.31311/ji.v5i1.2736>
- Aziz, F. A., Malik, R. F., Prasteyo, A. P. P., Komputer, F. I., & Sriwijaya, U. (2021). *Estimasi Posisi Objek Menggunakan Particle Swarm Optimization dan K – Nearest Neighbour*. 29–33.
- Choubey, D. K., Kumar, P., Tripathi, S., & Kumar, S. (2020). Performance evaluation of classification methods with PCA and PSO for diabetes. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 9(1). <https://doi.org/10.1007/s13721-019-0210-8>
- Dr. dr. Eva Decroli, S.-K. F. (2019). *Diabetes Melitus Tipe 2* (S. dr. Alexander Kam & S. dr. G. P. D. dr. A. R. dr. Yanne Pradwi Efendi (eds.)). Pusat Penerbitan Bagian Ilmu Penyakit Dalam Fakultas Kedokteran Universitas Andalas.
- Fernanda, S. I., Ratnawati, D. E., & Adikara, P. P. (2017). Identifikasi Penyakit Diabetes Mellitus Menggunakan Metode Modified K- Nearest Neighbor (MKNN). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 1(6), 507–513.
- Haerani et al. (2018). Komparasi Akurasi Metode Correlated Naive Bayes Classifier Dan Naive Bayes Classifier Untuk Diagnosis Penyakit Diabetes Hairani , Gibran Satya Nugraha , Mokhammad Nurkholis Abdillah , Muhammad Innuddin InfoTekJar (Jurnal Nasional Informatika dan Teknolog. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, 3(1), 6–11.
- Hasanuddin, H. (2016). Perbandingan Algoritma Knn Dan Knn-Pso Untuk Klasifikasi Tingkat Pengetahuan Ibu Dalam Pemberian Asi Eksklusif.

Technologia: Jurnal Ilmiah, 7(1), 34–40. <https://doi.org/10.31602/tji.v7i1.610>

Hermawan, Y. D. (2017). *Implementasi Algoritma K-Nearest Neighbors dengan Particle Swarm Optimization dalam Klasifikasi Trouble pada Base Transceiver Station (BTS)*.

KEMENKES RI. (2019). Hari Diabetes Sedunia Tahun 2018. *Pusat Data Dan Informasi Kementrian Kesehatan RI*, 1–8.

Khanesar, M. A., Teshnehlal, M., & Shoorehdeli, M. A. (n.d.). *A Novel Binary Particle Swarm Optimization*. 1(1).

Kurniawan, F., & Ivandari. (2017). Komparasi Algoritma Data Mining Untuk Klasifikasi Penyakit Kanker Payudara. *Jurnal Stmik*, XII(1), 1–8.

Liklikwatil, R. D., Noersasongko, E., & Supriyanto, C. (2018). *Optimasi K-Nearest Neighbor Dengan Particle Swarm Optimization Untuk Memprediksi Harga Komoditi Karet*. 7(2), 172–182.

Mahardika, K. W., Sari, Y. A., & Arwan, A. (2018). *Optimasi K-Nearest Neighbour Menggunakan Particle Swarm Optimization pada Sistem Pakar untuk Monitoring Pengendalian Hama pada Tanaman Jeruk*. 2(9), 3333–3344.

Maulida, A. (2020). Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes. *Indonesian Journal of Data and Science*, 1(2), 29–33.

Mirqotussa'adah, M., Muslim, M. A., Sugiharti, E., Prasetyo, B., & Alimah, S. (2017). Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa Diabetes. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 8(2), 135. <https://doi.org/10.24843/lkjiti.2017.v08.i02.p07>

Nasution, D. A., Khotimah, H. H., & Chamidah, N. (2019). Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Computer Engineering, Science and System Journal*, 4(1), 78.

<https://doi.org/10.24114/cess.v4i1.11458>

- Pradana, A. C., & Aditsania, A. (2018). *Implementasi Algoritma Binary Particle Swarm Optimization (BPSO) dan C4 . 5 Decision Tree untuk Deteksi Kanker Berdasarkan Klasifikasi Microarray Data*. 5(3), 7665–7682.
- SHAILENDRA, D. Y. (2017). *Seleksi Fitur Menggunakan Metode Hybrid Particle Swarm Optimization Dengan Operasi Local Search (Hpso-Ls) Untuk Klasifikasi Data*.
- Ulya, S., Soeleman, M. A., & Budiman, F. (2021). Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Prioritas Bantuan Pembangunan Desa. *Techno.Com*, 20(1), 83–96. <https://doi.org/10.33633/tc.v20i1.4215>
- Wijanarto, W., & Puspitasari, R. (2019). Optimasi Algoritma Klasifikasi Biner dengan Tuning Parameter pada Penyakit Diabetes Mellitus. *Eksplora Informatika*, 9(1), 50–59. <https://doi.org/10.30864/eksplora.v9i1.257>
- Yunita, F. (2016). *Sistem Klasifikasi Penyakit Diabetes Mellitus Menggunakan Metode K-Nearest Neighbor (K-Nn)*. 223–230.
- Yunus, W. (2018). *Algoritma K-Nearest Neighbor Berbasis Particle Swarm Optimization Untuk Prediksi Penyakit Ginjal Kronik*. 2(2), 51–55.
- Saxena, K., Khan, Z., & Singh, S. (2014). *Diagnosis of Diabetes Mellitus using K Nearest Neighbor Algorithm*. *International Journal of Computer Science Trends and Technology (IJCST)*, 2(4), 36–43.

LAMPIRAN

SAMPEL DATA

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 6 | 148 | 72 | 35 | 0 | 33.6 | 0.627 | 50 | 1 |
| 1 | 85 | 66 | 29 | 0 | 26.6 | 0.351 | 31 | 0 |
| 8 | 183 | 64 | 0 | 0 | 23.3 | 0.672 | 32 | 1 |
| 1 | 89 | 66 | 23 | 94 | 28.1 | 0.167 | 21 | 0 |
| 0 | 137 | 40 | 35 | 168 | 43.1 | 2.288 | 33 | 1 |
| 5 | 116 | 74 | 0 | 0 | 25.6 | 0.201 | 30 | 0 |
| 3 | 78 | 50 | 32 | 88 | 31 | 0.248 | 26 | 1 |
| 10 | 115 | 0 | 0 | 0 | 35.3 | 0.134 | 29 | 0 |
| 2 | 197 | 70 | 45 | 543 | 30.5 | 0.158 | 53 | 1 |
| 8 | 125 | 96 | 0 | 0 | 0 | 0.232 | 54 | 1 |
| 4 | 110 | 92 | 0 | 0 | 37.6 | 0.191 | 30 | 0 |
| 10 | 168 | 74 | 0 | 0 | 38 | 0.537 | 34 | 1 |
| 10 | 139 | 80 | 0 | 0 | 27.1 | 1.441 | 57 | 0 |
| 1 | 189 | 60 | 23 | 846 | 30.1 | 0.398 | 59 | 1 |
| 5 | 166 | 72 | 19 | 175 | 25.8 | 0.587 | 51 | 1 |
| 7 | 100 | 0 | 0 | 0 | 30 | 0.484 | 32 | 1 |
| 0 | 118 | 84 | 47 | 230 | 45.8 | 0.551 | 31 | 1 |
| 7 | 107 | 74 | 0 | 0 | 29.6 | 0.254 | 31 | 1 |
| 1 | 103 | 30 | 38 | 83 | 43.3 | 0.183 | 33 | 0 |
| 1 | 115 | 70 | 30 | 96 | 34.6 | 0.529 | 32 | 1 |
| 3 | 126 | 88 | 41 | 235 | 39.3 | 0.704 | 27 | 0 |
| 8 | 99 | 84 | 0 | 0 | 35.4 | 0.388 | 50 | 0 |
| 7 | 196 | 90 | 0 | 0 | 39.8 | 0.451 | 41 | 1 |
| 9 | 119 | 80 | 35 | 0 | 29 | 0.263 | 29 | 1 |
| 11 | 143 | 94 | 33 | 146 | 36.6 | 0.254 | 51 | 1 |
| 10 | 125 | 70 | 26 | 115 | 31.1 | 0.205 | 41 | 1 |
| 7 | 147 | 76 | 0 | 0 | 39.4 | 0.257 | 43 | 1 |
| 1 | 97 | 66 | 15 | 140 | 23.2 | 0.487 | 22 | 0 |
| 13 | 145 | 82 | 19 | 110 | 22.2 | 0.245 | 57 | 0 |
| 5 | 117 | 92 | 0 | 0 | 34.1 | 0.337 | 38 | 0 |
| 5 | 109 | 75 | 26 | 0 | 36 | 0.546 | 60 | 0 |
| 3 | 158 | 76 | 36 | 245 | 31.6 | 0.851 | 28 | 1 |
| 3 | 88 | 58 | 11 | 54 | 24.8 | 0.267 | 22 | 0 |
| 6 | 92 | 92 | 0 | 0 | 19.9 | 0.188 | 28 | 0 |
| 10 | 122 | 78 | 31 | 0 | 27.6 | 0.512 | 45 | 0 |
| 4 | 103 | 60 | 33 | 192 | 24 | 0.966 | 33 | 0 |
| 11 | 138 | 76 | 0 | 0 | 33.2 | 0.42 | 35 | 0 |
| 9 | 102 | 76 | 37 | 0 | 32.9 | 0.665 | 46 | 1 |
| 2 | 90 | 68 | 42 | 0 | 38.2 | 0.503 | 27 | 1 |
| 4 | 111 | 72 | 47 | 207 | 37.1 | 1.39 | 56 | 1 |
| 3 | 180 | 64 | 25 | 70 | 34 | 0.271 | 26 | 0 |
| 7 | 133 | 84 | 0 | 0 | 40.2 | 0.696 | 37 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 7 | 106 | 92 | 18 | 0 | 22.7 | 0.235 | 48 | 0 |
| 9 | 171 | 110 | 24 | 240 | 45.4 | 0.721 | 54 | 1 |
| 7 | 159 | 64 | 0 | 0 | 27.4 | 0.294 | 40 | 0 |
| 0 | 180 | 66 | 39 | 0 | 42 | 1.893 | 25 | 1 |
| 1 | 146 | 56 | 0 | 0 | 29.7 | 0.564 | 29 | 0 |
| 2 | 71 | 70 | 27 | 0 | 28 | 0.586 | 22 | 0 |
| 7 | 103 | 66 | 32 | 0 | 39.1 | 0.344 | 31 | 1 |
| 7 | 105 | 0 | 0 | 0 | 0 | 0.305 | 24 | 0 |
| 1 | 103 | 80 | 11 | 82 | 19.4 | 0.491 | 22 | 0 |
| 1 | 101 | 50 | 15 | 36 | 24.2 | 0.526 | 26 | 0 |
| 5 | 88 | 66 | 21 | 23 | 24.4 | 0.342 | 30 | 0 |
| 8 | 176 | 90 | 34 | 300 | 33.7 | 0.467 | 58 | 1 |
| 7 | 150 | 66 | 42 | 342 | 34.7 | 0.718 | 42 | 0 |
| 1 | 73 | 50 | 10 | 0 | 23 | 0.248 | 21 | 0 |
| 7 | 187 | 68 | 39 | 304 | 37.7 | 0.254 | 41 | 1 |
| 0 | 100 | 88 | 60 | 110 | 46.8 | 0.962 | 31 | 0 |
| 0 | 146 | 82 | 0 | 0 | 40.5 | 1.781 | 44 | 0 |
| 0 | 105 | 64 | 41 | 142 | 41.5 | 0.173 | 22 | 0 |
| 2 | 84 | 0 | 0 | 0 | 0 | 0.304 | 21 | 0 |
| 8 | 133 | 72 | 0 | 0 | 32.9 | 0.27 | 39 | 1 |
| 5 | 44 | 62 | 0 | 0 | 25 | 0.587 | 36 | 0 |
| 2 | 141 | 58 | 34 | 128 | 25.4 | 0.699 | 24 | 0 |
| 7 | 114 | 66 | 0 | 0 | 32.8 | 0.258 | 42 | 1 |
| 5 | 99 | 74 | 27 | 0 | 29 | 0.203 | 32 | 0 |
| 0 | 109 | 88 | 30 | 0 | 32.5 | 0.855 | 38 | 1 |
| 2 | 109 | 92 | 0 | 0 | 42.7 | 0.845 | 54 | 0 |
| 1 | 95 | 66 | 13 | 38 | 19.6 | 0.334 | 25 | 0 |
| 4 | 146 | 85 | 27 | 100 | 28.9 | 0.189 | 27 | 0 |
| 2 | 100 | 66 | 20 | 90 | 32.9 | 0.867 | 28 | 1 |
| 5 | 139 | 64 | 35 | 140 | 28.6 | 0.411 | 26 | 0 |
| 13 | 126 | 90 | 0 | 0 | 43.4 | 0.583 | 42 | 1 |
| 4 | 129 | 86 | 20 | 270 | 35.1 | 0.231 | 23 | 0 |
| 1 | 79 | 75 | 30 | 0 | 32 | 0.396 | 22 | 0 |
| 1 | 0 | 48 | 20 | 0 | 24.7 | 0.14 | 22 | 0 |
| 7 | 62 | 78 | 0 | 0 | 32.6 | 0.391 | 41 | 0 |
| 5 | 95 | 72 | 33 | 0 | 37.7 | 0.37 | 27 | 0 |
| 0 | 131 | 0 | 0 | 0 | 43.2 | 0.27 | 26 | 1 |
| 2 | 112 | 66 | 22 | 0 | 25 | 0.307 | 24 | 0 |
| 3 | 113 | 44 | 13 | 0 | 22.4 | 0.14 | 22 | 0 |
| 2 | 74 | 0 | 0 | 0 | 0 | 0.102 | 22 | 0 |
| 7 | 83 | 78 | 26 | 71 | 29.3 | 0.767 | 36 | 0 |
| 0 | 101 | 65 | 28 | 0 | 24.6 | 0.237 | 22 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 5 | 137 | 108 | 0 | 0 | 48.8 | 0.227 | 37 | 1 |
| 2 | 110 | 74 | 29 | 125 | 32.4 | 0.698 | 27 | 0 |
| 13 | 106 | 72 | 54 | 0 | 36.6 | 0.178 | 45 | 0 |
| 2 | 100 | 68 | 25 | 71 | 38.5 | 0.324 | 26 | 0 |
| 15 | 136 | 70 | 32 | 110 | 37.1 | 0.153 | 43 | 1 |
| 1 | 107 | 68 | 19 | 0 | 26.5 | 0.165 | 24 | 0 |
| 1 | 80 | 55 | 0 | 0 | 19.1 | 0.258 | 21 | 0 |
| 4 | 123 | 80 | 15 | 176 | 32 | 0.443 | 34 | 0 |
| 7 | 81 | 78 | 40 | 48 | 46.7 | 0.261 | 42 | 0 |
| 4 | 134 | 72 | 0 | 0 | 23.8 | 0.277 | 60 | 1 |
| 2 | 142 | 82 | 18 | 64 | 24.7 | 0.761 | 21 | 0 |
| 6 | 144 | 72 | 27 | 228 | 33.9 | 0.255 | 40 | 0 |
| 2 | 92 | 62 | 28 | 0 | 31.6 | 0.13 | 24 | 0 |
| 1 | 71 | 48 | 18 | 76 | 20.4 | 0.323 | 22 | 0 |
| 6 | 93 | 50 | 30 | 64 | 28.7 | 0.356 | 23 | 0 |
| 1 | 122 | 90 | 51 | 220 | 49.7 | 0.325 | 31 | 1 |
| 1 | 163 | 72 | 0 | 0 | 39 | 1.222 | 33 | 1 |
| 1 | 151 | 60 | 0 | 0 | 26.1 | 0.179 | 22 | 0 |
| 0 | 125 | 96 | 0 | 0 | 22.5 | 0.262 | 21 | 0 |
| 1 | 81 | 72 | 18 | 40 | 26.6 | 0.283 | 24 | 0 |
| 2 | 85 | 65 | 0 | 0 | 39.6 | 0.93 | 27 | 0 |
| 1 | 126 | 56 | 29 | 152 | 28.7 | 0.801 | 21 | 0 |
| 1 | 96 | 122 | 0 | 0 | 22.4 | 0.207 | 27 | 0 |
| 4 | 144 | 58 | 28 | 140 | 29.5 | 0.287 | 37 | 0 |
| 3 | 83 | 58 | 31 | 18 | 34.3 | 0.336 | 25 | 0 |
| 0 | 95 | 85 | 25 | 36 | 37.4 | 0.247 | 24 | 1 |
| 3 | 171 | 72 | 33 | 135 | 33.3 | 0.199 | 24 | 1 |
| 8 | 155 | 62 | 26 | 495 | 34 | 0.543 | 46 | 1 |
| 1 | 89 | 76 | 34 | 37 | 31.2 | 0.192 | 23 | 0 |
| 4 | 76 | 62 | 0 | 0 | 34 | 0.391 | 25 | 0 |
| 7 | 160 | 54 | 32 | 175 | 30.5 | 0.588 | 39 | 1 |
| 4 | 146 | 92 | 0 | 0 | 31.2 | 0.539 | 61 | 1 |
| 5 | 124 | 74 | 0 | 0 | 34 | 0.22 | 38 | 1 |
| 5 | 78 | 48 | 0 | 0 | 33.7 | 0.654 | 25 | 0 |
| 4 | 97 | 60 | 23 | 0 | 28.2 | 0.443 | 22 | 0 |
| 4 | 99 | 76 | 15 | 51 | 23.2 | 0.223 | 21 | 0 |
| 0 | 162 | 76 | 56 | 100 | 53.2 | 0.759 | 25 | 1 |
| 6 | 111 | 64 | 39 | 0 | 34.2 | 0.26 | 24 | 0 |
| 2 | 107 | 74 | 30 | 100 | 33.6 | 0.404 | 23 | 0 |
| 5 | 132 | 80 | 0 | 0 | 26.8 | 0.186 | 69 | 0 |
| 0 | 113 | 76 | 0 | 0 | 33.3 | 0.278 | 23 | 1 |
| 1 | 88 | 30 | 42 | 99 | 55 | 0.496 | 26 | 1 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 3 | 120 | 70 | 30 | 135 | 42.9 | 0.452 | 30 | 0 |
| 1 | 118 | 58 | 36 | 94 | 33.3 | 0.261 | 23 | 0 |
| 1 | 117 | 88 | 24 | 145 | 34.5 | 0.403 | 40 | 1 |
| 0 | 105 | 84 | 0 | 0 | 27.9 | 0.741 | 62 | 1 |
| 4 | 173 | 70 | 14 | 168 | 29.7 | 0.361 | 33 | 1 |
| 9 | 122 | 56 | 0 | 0 | 33.3 | 1.114 | 33 | 1 |
| 3 | 170 | 64 | 37 | 225 | 34.5 | 0.356 | 30 | 1 |
| 8 | 84 | 74 | 31 | 0 | 38.3 | 0.457 | 39 | 0 |
| 2 | 96 | 68 | 13 | 49 | 21.1 | 0.647 | 26 | 0 |
| 2 | 125 | 60 | 20 | 140 | 33.8 | 0.088 | 31 | 0 |
| 0 | 100 | 70 | 26 | 50 | 30.8 | 0.597 | 21 | 0 |
| 0 | 93 | 60 | 25 | 92 | 28.7 | 0.532 | 22 | 0 |
| 0 | 129 | 80 | 0 | 0 | 31.2 | 0.703 | 29 | 0 |
| 5 | 105 | 72 | 29 | 325 | 36.9 | 0.159 | 28 | 0 |
| 3 | 128 | 78 | 0 | 0 | 21.1 | 0.268 | 55 | 0 |
| 5 | 106 | 82 | 30 | 0 | 39.5 | 0.286 | 38 | 0 |
| 2 | 108 | 52 | 26 | 63 | 32.5 | 0.318 | 22 | 0 |
| 10 | 108 | 66 | 0 | 0 | 32.4 | 0.272 | 42 | 1 |
| 4 | 154 | 62 | 31 | 284 | 32.8 | 0.237 | 23 | 0 |
| 0 | 102 | 75 | 23 | 0 | 0 | 0.572 | 21 | 0 |
| 9 | 57 | 80 | 37 | 0 | 32.8 | 0.096 | 41 | 0 |
| 2 | 106 | 64 | 35 | 119 | 30.5 | 1.4 | 34 | 0 |
| 5 | 147 | 78 | 0 | 0 | 33.7 | 0.218 | 65 | 0 |
| 2 | 90 | 70 | 17 | 0 | 27.3 | 0.085 | 22 | 0 |
| 1 | 136 | 74 | 50 | 204 | 37.4 | 0.399 | 24 | 0 |
| 4 | 114 | 65 | 0 | 0 | 21.9 | 0.432 | 37 | 0 |
| 9 | 156 | 86 | 28 | 155 | 34.3 | 1.189 | 42 | 1 |
| 1 | 153 | 82 | 42 | 485 | 40.6 | 0.687 | 23 | 0 |
| 8 | 188 | 78 | 0 | 0 | 47.9 | 0.137 | 43 | 1 |
| 7 | 152 | 88 | 44 | 0 | 50 | 0.337 | 36 | 1 |
| 2 | 99 | 52 | 15 | 94 | 24.6 | 0.637 | 21 | 0 |
| 1 | 109 | 56 | 21 | 135 | 25.2 | 0.833 | 23 | 0 |
| 2 | 88 | 74 | 19 | 53 | 29 | 0.229 | 22 | 0 |
| 17 | 163 | 72 | 41 | 114 | 40.9 | 0.817 | 47 | 1 |
| 4 | 151 | 90 | 38 | 0 | 29.7 | 0.294 | 36 | 0 |
| 7 | 102 | 74 | 40 | 105 | 37.2 | 0.204 | 45 | 0 |
| 0 | 114 | 80 | 34 | 285 | 44.2 | 0.167 | 27 | 0 |
| 2 | 100 | 64 | 23 | 0 | 29.7 | 0.368 | 21 | 0 |
| 0 | 131 | 88 | 0 | 0 | 31.6 | 0.743 | 32 | 1 |
| 6 | 104 | 74 | 18 | 156 | 29.9 | 0.722 | 41 | 1 |
| 3 | 148 | 66 | 25 | 0 | 32.5 | 0.256 | 22 | 0 |
| 4 | 120 | 68 | 0 | 0 | 29.6 | 0.709 | 34 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 4 | 110 | 66 | 0 | 0 | 31.9 | 0.471 | 29 | 0 |
| 3 | 111 | 90 | 12 | 78 | 28.4 | 0.495 | 29 | 0 |
| 6 | 102 | 82 | 0 | 0 | 30.8 | 0.18 | 36 | 1 |
| 6 | 134 | 70 | 23 | 130 | 35.4 | 0.542 | 29 | 1 |
| 2 | 87 | 0 | 23 | 0 | 28.9 | 0.773 | 25 | 0 |
| 1 | 79 | 60 | 42 | 48 | 43.5 | 0.678 | 23 | 0 |
| 2 | 75 | 64 | 24 | 55 | 29.7 | 0.37 | 33 | 0 |
| 8 | 179 | 72 | 42 | 130 | 32.7 | 0.719 | 36 | 1 |
| 6 | 85 | 78 | 0 | 0 | 31.2 | 0.382 | 42 | 0 |
| 0 | 129 | 110 | 46 | 130 | 67.1 | 0.319 | 26 | 1 |
| 5 | 143 | 78 | 0 | 0 | 45 | 0.19 | 47 | 0 |
| 5 | 130 | 82 | 0 | 0 | 39.1 | 0.956 | 37 | 1 |
| 6 | 87 | 80 | 0 | 0 | 23.2 | 0.084 | 32 | 0 |
| 0 | 119 | 64 | 18 | 92 | 34.9 | 0.725 | 23 | 0 |
| 1 | 0 | 74 | 20 | 23 | 27.7 | 0.299 | 21 | 0 |
| 5 | 73 | 60 | 0 | 0 | 26.8 | 0.268 | 27 | 0 |
| 4 | 141 | 74 | 0 | 0 | 27.6 | 0.244 | 40 | 0 |
| 7 | 194 | 68 | 28 | 0 | 35.9 | 0.745 | 41 | 1 |
| 8 | 181 | 68 | 36 | 495 | 30.1 | 0.615 | 60 | 1 |
| 1 | 128 | 98 | 41 | 58 | 32 | 1.321 | 33 | 1 |
| 8 | 109 | 76 | 39 | 114 | 27.9 | 0.64 | 31 | 1 |
| 5 | 139 | 80 | 35 | 160 | 31.6 | 0.361 | 25 | 1 |
| 3 | 111 | 62 | 0 | 0 | 22.6 | 0.142 | 21 | 0 |
| 9 | 123 | 70 | 44 | 94 | 33.1 | 0.374 | 40 | 0 |
| 7 | 159 | 66 | 0 | 0 | 30.4 | 0.383 | 36 | 1 |
| 11 | 135 | 0 | 0 | 0 | 52.3 | 0.578 | 40 | 1 |
| 8 | 85 | 55 | 20 | 0 | 24.4 | 0.136 | 42 | 0 |
| 5 | 158 | 84 | 41 | 210 | 39.4 | 0.395 | 29 | 1 |
| 1 | 105 | 58 | 0 | 0 | 24.3 | 0.187 | 21 | 0 |
| 3 | 107 | 62 | 13 | 48 | 22.9 | 0.678 | 23 | 1 |
| 4 | 109 | 64 | 44 | 99 | 34.8 | 0.905 | 26 | 1 |
| 4 | 148 | 60 | 27 | 318 | 30.9 | 0.15 | 29 | 1 |
| 0 | 113 | 80 | 16 | 0 | 31 | 0.874 | 21 | 0 |
| 1 | 138 | 82 | 0 | 0 | 40.1 | 0.236 | 28 | 0 |
| 0 | 108 | 68 | 20 | 0 | 27.3 | 0.787 | 32 | 0 |
| 2 | 99 | 70 | 16 | 44 | 20.4 | 0.235 | 27 | 0 |
| 6 | 103 | 72 | 32 | 190 | 37.7 | 0.324 | 55 | 0 |
| 5 | 111 | 72 | 28 | 0 | 23.9 | 0.407 | 27 | 0 |
| 8 | 196 | 76 | 29 | 280 | 37.5 | 0.605 | 57 | 1 |
| 5 | 162 | 104 | 0 | 0 | 37.7 | 0.151 | 52 | 1 |
| 1 | 96 | 64 | 27 | 87 | 33.2 | 0.289 | 21 | 0 |
| 7 | 184 | 84 | 33 | 0 | 35.5 | 0.355 | 41 | 1 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 2 | 81 | 60 | 22 | 0 | 27.7 | 0.29 | 25 | 0 |
| 0 | 147 | 85 | 54 | 0 | 42.8 | 0.375 | 24 | 0 |
| 7 | 179 | 95 | 31 | 0 | 34.2 | 0.164 | 60 | 0 |
| 0 | 140 | 65 | 26 | 130 | 42.6 | 0.431 | 24 | 1 |
| 9 | 112 | 82 | 32 | 175 | 34.2 | 0.26 | 36 | 1 |
| 12 | 151 | 70 | 40 | 271 | 41.8 | 0.742 | 38 | 1 |
| 5 | 109 | 62 | 41 | 129 | 35.8 | 0.514 | 25 | 1 |
| 6 | 125 | 68 | 30 | 120 | 30 | 0.464 | 32 | 0 |
| 5 | 85 | 74 | 22 | 0 | 29 | 1.224 | 32 | 1 |
| 5 | 112 | 66 | 0 | 0 | 37.8 | 0.261 | 41 | 1 |
| 0 | 177 | 60 | 29 | 478 | 34.6 | 1.072 | 21 | 1 |
| 2 | 158 | 90 | 0 | 0 | 31.6 | 0.805 | 66 | 1 |
| 7 | 119 | 0 | 0 | 0 | 25.2 | 0.209 | 37 | 0 |
| 7 | 142 | 60 | 33 | 190 | 28.8 | 0.687 | 61 | 0 |
| 1 | 100 | 66 | 15 | 56 | 23.6 | 0.666 | 26 | 0 |
| 1 | 87 | 78 | 27 | 32 | 34.6 | 0.101 | 22 | 0 |
| 0 | 101 | 76 | 0 | 0 | 35.7 | 0.198 | 26 | 0 |
| 3 | 162 | 52 | 38 | 0 | 37.2 | 0.652 | 24 | 1 |
| 4 | 197 | 70 | 39 | 744 | 36.7 | 2.329 | 31 | 0 |
| 0 | 117 | 80 | 31 | 53 | 45.2 | 0.089 | 24 | 0 |
| 4 | 142 | 86 | 0 | 0 | 44 | 0.645 | 22 | 1 |
| 6 | 134 | 80 | 37 | 370 | 46.2 | 0.238 | 46 | 1 |
| 1 | 79 | 80 | 25 | 37 | 25.4 | 0.583 | 22 | 0 |
| 4 | 122 | 68 | 0 | 0 | 35 | 0.394 | 29 | 0 |
| 3 | 74 | 68 | 28 | 45 | 29.7 | 0.293 | 23 | 0 |
| 4 | 171 | 72 | 0 | 0 | 43.6 | 0.479 | 26 | 1 |
| 7 | 181 | 84 | 21 | 192 | 35.9 | 0.586 | 51 | 1 |
| 0 | 179 | 90 | 27 | 0 | 44.1 | 0.686 | 23 | 1 |
| 9 | 164 | 84 | 21 | 0 | 30.8 | 0.831 | 32 | 1 |
| 0 | 104 | 76 | 0 | 0 | 18.4 | 0.582 | 27 | 0 |
| 1 | 91 | 64 | 24 | 0 | 29.2 | 0.192 | 21 | 0 |
| 4 | 91 | 70 | 32 | 88 | 33.1 | 0.446 | 22 | 0 |
| 3 | 139 | 54 | 0 | 0 | 25.6 | 0.402 | 22 | 1 |
| 6 | 119 | 50 | 22 | 176 | 27.1 | 1.318 | 33 | 1 |
| 2 | 146 | 76 | 35 | 194 | 38.2 | 0.329 | 29 | 0 |
| 9 | 184 | 85 | 15 | 0 | 30 | 1.213 | 49 | 1 |
| 10 | 122 | 68 | 0 | 0 | 31.2 | 0.258 | 41 | 0 |
| 0 | 165 | 90 | 33 | 680 | 52.3 | 0.427 | 23 | 0 |
| 9 | 124 | 70 | 33 | 402 | 35.4 | 0.282 | 34 | 0 |
| 1 | 111 | 86 | 19 | 0 | 30.1 | 0.143 | 23 | 0 |
| 9 | 106 | 52 | 0 | 0 | 31.2 | 0.38 | 42 | 0 |
| 2 | 129 | 84 | 0 | 0 | 28 | 0.284 | 27 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 2 | 90 | 80 | 14 | 55 | 24.4 | 0.249 | 24 | 0 |
| 0 | 86 | 68 | 32 | 0 | 35.8 | 0.238 | 25 | 0 |
| 12 | 92 | 62 | 7 | 258 | 27.6 | 0.926 | 44 | 1 |
| 1 | 113 | 64 | 35 | 0 | 33.6 | 0.543 | 21 | 1 |
| 3 | 111 | 56 | 39 | 0 | 30.1 | 0.557 | 30 | 0 |
| 2 | 114 | 68 | 22 | 0 | 28.7 | 0.092 | 25 | 0 |
| 1 | 193 | 50 | 16 | 375 | 25.9 | 0.655 | 24 | 0 |
| 11 | 155 | 76 | 28 | 150 | 33.3 | 1.353 | 51 | 1 |
| 3 | 191 | 68 | 15 | 130 | 30.9 | 0.299 | 34 | 0 |
| 3 | 141 | 0 | 0 | 0 | 30 | 0.761 | 27 | 1 |
| 4 | 95 | 70 | 32 | 0 | 32.1 | 0.612 | 24 | 0 |
| 3 | 142 | 80 | 15 | 0 | 32.4 | 0.2 | 63 | 0 |
| 4 | 123 | 62 | 0 | 0 | 32 | 0.226 | 35 | 1 |
| 5 | 96 | 74 | 18 | 67 | 33.6 | 0.997 | 43 | 0 |
| 0 | 138 | 0 | 0 | 0 | 36.3 | 0.933 | 25 | 1 |
| 2 | 128 | 64 | 42 | 0 | 40 | 1.101 | 24 | 0 |
| 0 | 102 | 52 | 0 | 0 | 25.1 | 0.078 | 21 | 0 |
| 2 | 146 | 0 | 0 | 0 | 27.5 | 0.24 | 28 | 1 |
| 10 | 101 | 86 | 37 | 0 | 45.6 | 1.136 | 38 | 1 |
| 2 | 108 | 62 | 32 | 56 | 25.2 | 0.128 | 21 | 0 |
| 3 | 122 | 78 | 0 | 0 | 23 | 0.254 | 40 | 0 |
| 1 | 71 | 78 | 50 | 45 | 33.2 | 0.422 | 21 | 0 |
| 13 | 106 | 70 | 0 | 0 | 34.2 | 0.251 | 52 | 0 |
| 2 | 100 | 70 | 52 | 57 | 40.5 | 0.677 | 25 | 0 |
| 7 | 106 | 60 | 24 | 0 | 26.5 | 0.296 | 29 | 1 |
| 0 | 104 | 64 | 23 | 116 | 27.8 | 0.454 | 23 | 0 |
| 5 | 114 | 74 | 0 | 0 | 24.9 | 0.744 | 57 | 0 |
| 2 | 108 | 62 | 10 | 278 | 25.3 | 0.881 | 22 | 0 |
| 0 | 146 | 70 | 0 | 0 | 37.9 | 0.334 | 28 | 1 |
| 10 | 129 | 76 | 28 | 122 | 35.9 | 0.28 | 39 | 0 |
| 7 | 133 | 88 | 15 | 155 | 32.4 | 0.262 | 37 | 0 |
| 7 | 161 | 86 | 0 | 0 | 30.4 | 0.165 | 47 | 1 |
| 2 | 108 | 80 | 0 | 0 | 27 | 0.259 | 52 | 1 |
| 7 | 136 | 74 | 26 | 135 | 26 | 0.647 | 51 | 0 |
| 5 | 155 | 84 | 44 | 545 | 38.7 | 0.619 | 34 | 0 |
| 1 | 119 | 86 | 39 | 220 | 45.6 | 0.808 | 29 | 1 |
| 4 | 96 | 56 | 17 | 49 | 20.8 | 0.34 | 26 | 0 |
| 5 | 108 | 72 | 43 | 75 | 36.1 | 0.263 | 33 | 0 |
| 0 | 78 | 88 | 29 | 40 | 36.9 | 0.434 | 21 | 0 |
| 0 | 107 | 62 | 30 | 74 | 36.6 | 0.757 | 25 | 1 |
| 2 | 128 | 78 | 37 | 182 | 43.3 | 1.224 | 31 | 1 |
| 1 | 128 | 48 | 45 | 194 | 40.5 | 0.613 | 24 | 1 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 0 | 161 | 50 | 0 | 0 | 21.9 | 0.254 | 65 | 0 |
| 6 | 151 | 62 | 31 | 120 | 35.5 | 0.692 | 28 | 0 |
| 2 | 146 | 70 | 38 | 360 | 28 | 0.337 | 29 | 1 |
| 0 | 126 | 84 | 29 | 215 | 30.7 | 0.52 | 24 | 0 |
| 14 | 100 | 78 | 25 | 184 | 36.6 | 0.412 | 46 | 1 |
| 8 | 112 | 72 | 0 | 0 | 23.6 | 0.84 | 58 | 0 |
| 0 | 167 | 0 | 0 | 0 | 32.3 | 0.839 | 30 | 1 |
| 2 | 144 | 58 | 33 | 135 | 31.6 | 0.422 | 25 | 1 |
| 5 | 77 | 82 | 41 | 42 | 35.8 | 0.156 | 35 | 0 |
| 5 | 115 | 98 | 0 | 0 | 52.9 | 0.209 | 28 | 1 |
| 3 | 150 | 76 | 0 | 0 | 21 | 0.207 | 37 | 0 |
| 2 | 120 | 76 | 37 | 105 | 39.7 | 0.215 | 29 | 0 |
| 10 | 161 | 68 | 23 | 132 | 25.5 | 0.326 | 47 | 1 |
| 0 | 137 | 68 | 14 | 148 | 24.8 | 0.143 | 21 | 0 |
| 0 | 128 | 68 | 19 | 180 | 30.5 | 1.391 | 25 | 1 |
| 2 | 124 | 68 | 28 | 205 | 32.9 | 0.875 | 30 | 1 |
| 6 | 80 | 66 | 30 | 0 | 26.2 | 0.313 | 41 | 0 |
| 0 | 106 | 70 | 37 | 148 | 39.4 | 0.605 | 22 | 0 |
| 2 | 155 | 74 | 17 | 96 | 26.6 | 0.433 | 27 | 1 |
| 3 | 113 | 50 | 10 | 85 | 29.5 | 0.626 | 25 | 0 |
| 7 | 109 | 80 | 31 | 0 | 35.9 | 1.127 | 43 | 1 |
| 2 | 112 | 68 | 22 | 94 | 34.1 | 0.315 | 26 | 0 |
| 3 | 99 | 80 | 11 | 64 | 19.3 | 0.284 | 30 | 0 |
| 3 | 182 | 74 | 0 | 0 | 30.5 | 0.345 | 29 | 1 |
| 3 | 115 | 66 | 39 | 140 | 38.1 | 0.15 | 28 | 0 |
| 6 | 194 | 78 | 0 | 0 | 23.5 | 0.129 | 59 | 1 |
| 4 | 129 | 60 | 12 | 231 | 27.5 | 0.527 | 31 | 0 |
| 3 | 112 | 74 | 30 | 0 | 31.6 | 0.197 | 25 | 1 |
| 0 | 124 | 70 | 20 | 0 | 27.4 | 0.254 | 36 | 1 |
| 13 | 152 | 90 | 33 | 29 | 26.8 | 0.731 | 43 | 1 |
| 2 | 112 | 75 | 32 | 0 | 35.7 | 0.148 | 21 | 0 |
| 1 | 157 | 72 | 21 | 168 | 25.6 | 0.123 | 24 | 0 |
| 1 | 122 | 64 | 32 | 156 | 35.1 | 0.692 | 30 | 1 |
| 10 | 179 | 70 | 0 | 0 | 35.1 | 0.2 | 37 | 0 |
| 2 | 102 | 86 | 36 | 120 | 45.5 | 0.127 | 23 | 1 |
| 6 | 105 | 70 | 32 | 68 | 30.8 | 0.122 | 37 | 0 |
| 8 | 118 | 72 | 19 | 0 | 23.1 | 1.476 | 46 | 0 |
| 2 | 87 | 58 | 16 | 52 | 32.7 | 0.166 | 25 | 0 |
| 1 | 180 | 0 | 0 | 0 | 43.3 | 0.282 | 41 | 1 |
| 12 | 106 | 80 | 0 | 0 | 23.6 | 0.137 | 44 | 0 |
| 1 | 95 | 60 | 18 | 58 | 23.9 | 0.26 | 22 | 0 |
| 0 | 165 | 76 | 43 | 255 | 47.9 | 0.259 | 26 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 0 | 117 | 0 | 0 | 0 | 33.8 | 0.932 | 44 | 0 |
| 5 | 115 | 76 | 0 | 0 | 31.2 | 0.343 | 44 | 1 |
| 9 | 152 | 78 | 34 | 171 | 34.2 | 0.893 | 33 | 1 |
| 7 | 178 | 84 | 0 | 0 | 39.9 | 0.331 | 41 | 1 |
| 1 | 130 | 70 | 13 | 105 | 25.9 | 0.472 | 22 | 0 |
| 1 | 95 | 74 | 21 | 73 | 25.9 | 0.673 | 36 | 0 |
| 1 | 0 | 68 | 35 | 0 | 32 | 0.389 | 22 | 0 |
| 5 | 122 | 86 | 0 | 0 | 34.7 | 0.29 | 33 | 0 |
| 8 | 95 | 72 | 0 | 0 | 36.8 | 0.485 | 57 | 0 |
| 8 | 126 | 88 | 36 | 108 | 38.5 | 0.349 | 49 | 0 |
| 1 | 139 | 46 | 19 | 83 | 28.7 | 0.654 | 22 | 0 |
| 3 | 116 | 0 | 0 | 0 | 23.5 | 0.187 | 23 | 0 |
| 3 | 99 | 62 | 19 | 74 | 21.8 | 0.279 | 26 | 0 |
| 5 | 0 | 80 | 32 | 0 | 41 | 0.346 | 37 | 1 |
| 4 | 92 | 80 | 0 | 0 | 42.2 | 0.237 | 29 | 0 |
| 4 | 137 | 84 | 0 | 0 | 31.2 | 0.252 | 30 | 0 |
| 3 | 61 | 82 | 28 | 0 | 34.4 | 0.243 | 46 | 0 |
| 1 | 90 | 62 | 12 | 43 | 27.2 | 0.58 | 24 | 0 |
| 3 | 90 | 78 | 0 | 0 | 42.7 | 0.559 | 21 | 0 |
| 9 | 165 | 88 | 0 | 0 | 30.4 | 0.302 | 49 | 1 |
| 1 | 125 | 50 | 40 | 167 | 33.3 | 0.962 | 28 | 1 |
| 13 | 129 | 0 | 30 | 0 | 39.9 | 0.569 | 44 | 1 |
| 12 | 88 | 74 | 40 | 54 | 35.3 | 0.378 | 48 | 0 |
| 1 | 196 | 76 | 36 | 249 | 36.5 | 0.875 | 29 | 1 |
| 5 | 189 | 64 | 33 | 325 | 31.2 | 0.583 | 29 | 1 |
| 5 | 158 | 70 | 0 | 0 | 29.8 | 0.207 | 63 | 0 |
| 5 | 103 | 108 | 37 | 0 | 39.2 | 0.305 | 65 | 0 |
| 4 | 146 | 78 | 0 | 0 | 38.5 | 0.52 | 67 | 1 |
| 4 | 147 | 74 | 25 | 293 | 34.9 | 0.385 | 30 | 0 |
| 5 | 99 | 54 | 28 | 83 | 34 | 0.499 | 30 | 0 |
| 6 | 124 | 72 | 0 | 0 | 27.6 | 0.368 | 29 | 1 |
| 0 | 101 | 64 | 17 | 0 | 21 | 0.252 | 21 | 0 |
| 3 | 81 | 86 | 16 | 66 | 27.5 | 0.306 | 22 | 0 |
| 1 | 133 | 102 | 28 | 140 | 32.8 | 0.234 | 45 | 1 |
| 3 | 173 | 82 | 48 | 465 | 38.4 | 2.137 | 25 | 1 |
| 0 | 118 | 64 | 23 | 89 | 0 | 1.731 | 21 | 0 |
| 0 | 84 | 64 | 22 | 66 | 35.8 | 0.545 | 21 | 0 |
| 2 | 105 | 58 | 40 | 94 | 34.9 | 0.225 | 25 | 0 |
| 2 | 122 | 52 | 43 | 158 | 36.2 | 0.816 | 28 | 0 |
| 12 | 140 | 82 | 43 | 325 | 39.2 | 0.528 | 58 | 1 |
| 0 | 98 | 82 | 15 | 84 | 25.2 | 0.299 | 22 | 0 |
| 1 | 87 | 60 | 37 | 75 | 37.2 | 0.509 | 22 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 4 | 156 | 75 | 0 | 0 | 48.3 | 0.238 | 32 | 1 |
| 0 | 93 | 100 | 39 | 72 | 43.4 | 1.021 | 35 | 0 |
| 1 | 107 | 72 | 30 | 82 | 30.8 | 0.821 | 24 | 0 |
| 0 | 105 | 68 | 22 | 0 | 20 | 0.236 | 22 | 0 |
| 1 | 109 | 60 | 8 | 182 | 25.4 | 0.947 | 21 | 0 |
| 1 | 90 | 62 | 18 | 59 | 25.1 | 1.268 | 25 | 0 |
| 1 | 125 | 70 | 24 | 110 | 24.3 | 0.221 | 25 | 0 |
| 1 | 119 | 54 | 13 | 50 | 22.3 | 0.205 | 24 | 0 |
| 5 | 116 | 74 | 29 | 0 | 32.3 | 0.66 | 35 | 1 |
| 8 | 105 | 100 | 36 | 0 | 43.3 | 0.239 | 45 | 1 |
| 5 | 144 | 82 | 26 | 285 | 32 | 0.452 | 58 | 1 |
| 3 | 100 | 68 | 23 | 81 | 31.6 | 0.949 | 28 | 0 |
| 1 | 100 | 66 | 29 | 196 | 32 | 0.444 | 42 | 0 |
| 5 | 166 | 76 | 0 | 0 | 45.7 | 0.34 | 27 | 1 |
| 1 | 131 | 64 | 14 | 415 | 23.7 | 0.389 | 21 | 0 |
| 4 | 116 | 72 | 12 | 87 | 22.1 | 0.463 | 37 | 0 |
| 4 | 158 | 78 | 0 | 0 | 32.9 | 0.803 | 31 | 1 |
| 2 | 127 | 58 | 24 | 275 | 27.7 | 1.6 | 25 | 0 |
| 3 | 96 | 56 | 34 | 115 | 24.7 | 0.944 | 39 | 0 |
| 0 | 131 | 66 | 40 | 0 | 34.3 | 0.196 | 22 | 1 |
| 3 | 82 | 70 | 0 | 0 | 21.1 | 0.389 | 25 | 0 |
| 3 | 193 | 70 | 31 | 0 | 34.9 | 0.241 | 25 | 1 |
| 4 | 95 | 64 | 0 | 0 | 32 | 0.161 | 31 | 1 |
| 6 | 137 | 61 | 0 | 0 | 24.2 | 0.151 | 55 | 0 |
| 5 | 136 | 84 | 41 | 88 | 35 | 0.286 | 35 | 1 |
| 9 | 72 | 78 | 25 | 0 | 31.6 | 0.28 | 38 | 0 |
| 5 | 168 | 64 | 0 | 0 | 32.9 | 0.135 | 41 | 1 |
| 2 | 123 | 48 | 32 | 165 | 42.1 | 0.52 | 26 | 0 |
| 4 | 115 | 72 | 0 | 0 | 28.9 | 0.376 | 46 | 1 |
| 0 | 101 | 62 | 0 | 0 | 21.9 | 0.336 | 25 | 0 |
| 8 | 197 | 74 | 0 | 0 | 25.9 | 1.191 | 39 | 1 |
| 1 | 172 | 68 | 49 | 579 | 42.4 | 0.702 | 28 | 1 |
| 6 | 102 | 90 | 39 | 0 | 35.7 | 0.674 | 28 | 0 |
| 1 | 112 | 72 | 30 | 176 | 34.4 | 0.528 | 25 | 0 |
| 1 | 143 | 84 | 23 | 310 | 42.4 | 1.076 | 22 | 0 |
| 1 | 143 | 74 | 22 | 61 | 26.2 | 0.256 | 21 | 0 |
| 0 | 138 | 60 | 35 | 167 | 34.6 | 0.534 | 21 | 1 |
| 3 | 173 | 84 | 33 | 474 | 35.7 | 0.258 | 22 | 1 |
| 1 | 97 | 68 | 21 | 0 | 27.2 | 1.095 | 22 | 0 |
| 4 | 144 | 82 | 32 | 0 | 38.5 | 0.554 | 37 | 1 |
| 1 | 83 | 68 | 0 | 0 | 18.2 | 0.624 | 27 | 0 |
| 3 | 129 | 64 | 29 | 115 | 26.4 | 0.219 | 28 | 1 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 1 | 119 | 88 | 41 | 170 | 45.3 | 0.507 | 26 | 0 |
| 2 | 94 | 68 | 18 | 76 | 26 | 0.561 | 21 | 0 |
| 0 | 102 | 64 | 46 | 78 | 40.6 | 0.496 | 21 | 0 |
| 2 | 115 | 64 | 22 | 0 | 30.8 | 0.421 | 21 | 0 |
| 8 | 151 | 78 | 32 | 210 | 42.9 | 0.516 | 36 | 1 |
| 4 | 184 | 78 | 39 | 277 | 37 | 0.264 | 31 | 1 |
| 0 | 94 | 0 | 0 | 0 | 0 | 0.256 | 25 | 0 |
| 1 | 181 | 64 | 30 | 180 | 34.1 | 0.328 | 38 | 1 |
| 0 | 135 | 94 | 46 | 145 | 40.6 | 0.284 | 26 | 0 |
| 1 | 95 | 82 | 25 | 180 | 35 | 0.233 | 43 | 1 |
| 2 | 99 | 0 | 0 | 0 | 22.2 | 0.108 | 23 | 0 |
| 3 | 89 | 74 | 16 | 85 | 30.4 | 0.551 | 38 | 0 |
| 1 | 80 | 74 | 11 | 60 | 30 | 0.527 | 22 | 0 |
| 2 | 139 | 75 | 0 | 0 | 25.6 | 0.167 | 29 | 0 |
| 1 | 90 | 68 | 8 | 0 | 24.5 | 1.138 | 36 | 0 |
| 0 | 141 | 0 | 0 | 0 | 42.4 | 0.205 | 29 | 1 |
| 12 | 140 | 85 | 33 | 0 | 37.4 | 0.244 | 41 | 0 |
| 5 | 147 | 75 | 0 | 0 | 29.9 | 0.434 | 28 | 0 |
| 1 | 97 | 70 | 15 | 0 | 18.2 | 0.147 | 21 | 0 |
| 6 | 107 | 88 | 0 | 0 | 36.8 | 0.727 | 31 | 0 |
| 0 | 189 | 104 | 25 | 0 | 34.3 | 0.435 | 41 | 1 |
| 2 | 83 | 66 | 23 | 50 | 32.2 | 0.497 | 22 | 0 |
| 4 | 117 | 64 | 27 | 120 | 33.2 | 0.23 | 24 | 0 |
| 8 | 108 | 70 | 0 | 0 | 30.5 | 0.955 | 33 | 1 |
| 4 | 117 | 62 | 12 | 0 | 29.7 | 0.38 | 30 | 1 |
| 0 | 180 | 78 | 63 | 14 | 59.4 | 2.42 | 25 | 1 |
| 1 | 100 | 72 | 12 | 70 | 25.3 | 0.658 | 28 | 0 |
| 0 | 95 | 80 | 45 | 92 | 36.5 | 0.33 | 26 | 0 |
| 0 | 104 | 64 | 37 | 64 | 33.6 | 0.51 | 22 | 1 |
| 0 | 120 | 74 | 18 | 63 | 30.5 | 0.285 | 26 | 0 |
| 1 | 82 | 64 | 13 | 95 | 21.2 | 0.415 | 23 | 0 |
| 2 | 134 | 70 | 0 | 0 | 28.9 | 0.542 | 23 | 1 |
| 0 | 91 | 68 | 32 | 210 | 39.9 | 0.381 | 25 | 0 |
| 2 | 119 | 0 | 0 | 0 | 19.6 | 0.832 | 72 | 0 |
| 2 | 100 | 54 | 28 | 105 | 37.8 | 0.498 | 24 | 0 |
| 14 | 175 | 62 | 30 | 0 | 33.6 | 0.212 | 38 | 1 |
| 1 | 135 | 54 | 0 | 0 | 26.7 | 0.687 | 62 | 0 |
| 5 | 86 | 68 | 28 | 71 | 30.2 | 0.364 | 24 | 0 |
| 10 | 148 | 84 | 48 | 237 | 37.6 | 1.001 | 51 | 1 |
| 9 | 134 | 74 | 33 | 60 | 25.9 | 0.46 | 81 | 0 |
| 9 | 120 | 72 | 22 | 56 | 20.8 | 0.733 | 48 | 0 |
| 1 | 71 | 62 | 0 | 0 | 21.8 | 0.416 | 26 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 8 | 74 | 70 | 40 | 49 | 35.3 | 0.705 | 39 | 0 |
| 5 | 88 | 78 | 30 | 0 | 27.6 | 0.258 | 37 | 0 |
| 10 | 115 | 98 | 0 | 0 | 24 | 1.022 | 34 | 0 |
| 0 | 124 | 56 | 13 | 105 | 21.8 | 0.452 | 21 | 0 |
| 0 | 74 | 52 | 10 | 36 | 27.8 | 0.269 | 22 | 0 |
| 0 | 97 | 64 | 36 | 100 | 36.8 | 0.6 | 25 | 0 |
| 8 | 120 | 0 | 0 | 0 | 30 | 0.183 | 38 | 1 |
| 6 | 154 | 78 | 41 | 140 | 46.1 | 0.571 | 27 | 0 |
| 1 | 144 | 82 | 40 | 0 | 41.3 | 0.607 | 28 | 0 |
| 0 | 137 | 70 | 38 | 0 | 33.2 | 0.17 | 22 | 0 |
| 0 | 119 | 66 | 27 | 0 | 38.8 | 0.259 | 22 | 0 |
| 7 | 136 | 90 | 0 | 0 | 29.9 | 0.21 | 50 | 0 |
| 4 | 114 | 64 | 0 | 0 | 28.9 | 0.126 | 24 | 0 |
| 0 | 137 | 84 | 27 | 0 | 27.3 | 0.231 | 59 | 0 |
| 2 | 105 | 80 | 45 | 191 | 33.7 | 0.711 | 29 | 1 |
| 7 | 114 | 76 | 17 | 110 | 23.8 | 0.466 | 31 | 0 |
| 8 | 126 | 74 | 38 | 75 | 25.9 | 0.162 | 39 | 0 |
| 4 | 132 | 86 | 31 | 0 | 28 | 0.419 | 63 | 0 |
| 3 | 158 | 70 | 30 | 328 | 35.5 | 0.344 | 35 | 1 |
| 0 | 123 | 88 | 37 | 0 | 35.2 | 0.197 | 29 | 0 |
| 4 | 85 | 58 | 22 | 49 | 27.8 | 0.306 | 28 | 0 |
| 0 | 84 | 82 | 31 | 125 | 38.2 | 0.233 | 23 | 0 |
| 0 | 145 | 0 | 0 | 0 | 44.2 | 0.63 | 31 | 1 |
| 0 | 135 | 68 | 42 | 250 | 42.3 | 0.365 | 24 | 1 |
| 1 | 139 | 62 | 41 | 480 | 40.7 | 0.536 | 21 | 0 |
| 0 | 173 | 78 | 32 | 265 | 46.5 | 1.159 | 58 | 0 |
| 4 | 99 | 72 | 17 | 0 | 25.6 | 0.294 | 28 | 0 |
| 8 | 194 | 80 | 0 | 0 | 26.1 | 0.551 | 67 | 0 |
| 2 | 83 | 65 | 28 | 66 | 36.8 | 0.629 | 24 | 0 |
| 2 | 89 | 90 | 30 | 0 | 33.5 | 0.292 | 42 | 0 |
| 4 | 99 | 68 | 38 | 0 | 32.8 | 0.145 | 33 | 0 |
| 4 | 125 | 70 | 18 | 122 | 28.9 | 1.144 | 45 | 1 |
| 3 | 80 | 0 | 0 | 0 | 0 | 0.174 | 22 | 0 |
| 6 | 166 | 74 | 0 | 0 | 26.6 | 0.304 | 66 | 0 |
| 5 | 110 | 68 | 0 | 0 | 26 | 0.292 | 30 | 0 |
| 2 | 81 | 72 | 15 | 76 | 30.1 | 0.547 | 25 | 0 |
| 7 | 195 | 70 | 33 | 145 | 25.1 | 0.163 | 55 | 1 |
| 6 | 154 | 74 | 32 | 193 | 29.3 | 0.839 | 39 | 0 |
| 2 | 117 | 90 | 19 | 71 | 25.2 | 0.313 | 21 | 0 |
| 3 | 84 | 72 | 32 | 0 | 37.2 | 0.267 | 28 | 0 |
| 6 | 0 | 68 | 41 | 0 | 39 | 0.727 | 41 | 1 |
| 7 | 94 | 64 | 25 | 79 | 33.3 | 0.738 | 41 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 3 | 96 | 78 | 39 | 0 | 37.3 | 0.238 | 40 | 0 |
| 10 | 75 | 82 | 0 | 0 | 33.3 | 0.263 | 38 | 0 |
| 0 | 180 | 90 | 26 | 90 | 36.5 | 0.314 | 35 | 1 |
| 1 | 130 | 60 | 23 | 170 | 28.6 | 0.692 | 21 | 0 |
| 2 | 84 | 50 | 23 | 76 | 30.4 | 0.968 | 21 | 0 |
| 8 | 120 | 78 | 0 | 0 | 25 | 0.409 | 64 | 0 |
| 12 | 84 | 72 | 31 | 0 | 29.7 | 0.297 | 46 | 1 |
| 0 | 139 | 62 | 17 | 210 | 22.1 | 0.207 | 21 | 0 |
| 9 | 91 | 68 | 0 | 0 | 24.2 | 0.2 | 58 | 0 |
| 2 | 91 | 62 | 0 | 0 | 27.3 | 0.525 | 22 | 0 |
| 3 | 99 | 54 | 19 | 86 | 25.6 | 0.154 | 24 | 0 |
| 3 | 163 | 70 | 18 | 105 | 31.6 | 0.268 | 28 | 1 |
| 9 | 145 | 88 | 34 | 165 | 30.3 | 0.771 | 53 | 1 |
| 7 | 125 | 86 | 0 | 0 | 37.6 | 0.304 | 51 | 0 |
| 13 | 76 | 60 | 0 | 0 | 32.8 | 0.18 | 41 | 0 |
| 6 | 129 | 90 | 7 | 326 | 19.6 | 0.582 | 60 | 0 |
| 2 | 68 | 70 | 32 | 66 | 25 | 0.187 | 25 | 0 |
| 3 | 124 | 80 | 33 | 130 | 33.2 | 0.305 | 26 | 0 |
| 6 | 114 | 0 | 0 | 0 | 0 | 0.189 | 26 | 0 |
| 9 | 130 | 70 | 0 | 0 | 34.2 | 0.652 | 45 | 1 |
| 3 | 125 | 58 | 0 | 0 | 31.6 | 0.151 | 24 | 0 |
| 3 | 87 | 60 | 18 | 0 | 21.8 | 0.444 | 21 | 0 |
| 1 | 97 | 64 | 19 | 82 | 18.2 | 0.299 | 21 | 0 |
| 3 | 116 | 74 | 15 | 105 | 26.3 | 0.107 | 24 | 0 |
| 0 | 117 | 66 | 31 | 188 | 30.8 | 0.493 | 22 | 0 |
| 0 | 111 | 65 | 0 | 0 | 24.6 | 0.66 | 31 | 0 |
| 2 | 122 | 60 | 18 | 106 | 29.8 | 0.717 | 22 | 0 |
| 0 | 107 | 76 | 0 | 0 | 45.3 | 0.686 | 24 | 0 |
| 1 | 86 | 66 | 52 | 65 | 41.3 | 0.917 | 29 | 0 |
| 6 | 91 | 0 | 0 | 0 | 29.8 | 0.501 | 31 | 0 |
| 1 | 77 | 56 | 30 | 56 | 33.3 | 1.251 | 24 | 0 |
| 4 | 132 | 0 | 0 | 0 | 32.9 | 0.302 | 23 | 1 |
| 0 | 105 | 90 | 0 | 0 | 29.6 | 0.197 | 46 | 0 |
| 0 | 57 | 60 | 0 | 0 | 21.7 | 0.735 | 67 | 0 |
| 0 | 127 | 80 | 37 | 210 | 36.3 | 0.804 | 23 | 0 |
| 3 | 129 | 92 | 49 | 155 | 36.4 | 0.968 | 32 | 1 |
| 8 | 100 | 74 | 40 | 215 | 39.4 | 0.661 | 43 | 1 |
| 3 | 128 | 72 | 25 | 190 | 32.4 | 0.549 | 27 | 1 |
| 10 | 90 | 85 | 32 | 0 | 34.9 | 0.825 | 56 | 1 |
| 4 | 84 | 90 | 23 | 56 | 39.5 | 0.159 | 25 | 0 |
| 1 | 88 | 78 | 29 | 76 | 32 | 0.365 | 29 | 0 |
| 8 | 186 | 90 | 35 | 225 | 34.5 | 0.423 | 37 | 1 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 5 | 187 | 76 | 27 | 207 | 43.6 | 1.034 | 53 | 1 |
| 4 | 131 | 68 | 21 | 166 | 33.1 | 0.16 | 28 | 0 |
| 1 | 164 | 82 | 43 | 67 | 32.8 | 0.341 | 50 | 0 |
| 4 | 189 | 110 | 31 | 0 | 28.5 | 0.68 | 37 | 0 |
| 1 | 116 | 70 | 28 | 0 | 27.4 | 0.204 | 21 | 0 |
| 3 | 84 | 68 | 30 | 106 | 31.9 | 0.591 | 25 | 0 |
| 6 | 114 | 88 | 0 | 0 | 27.8 | 0.247 | 66 | 0 |
| 1 | 88 | 62 | 24 | 44 | 29.9 | 0.422 | 23 | 0 |
| 1 | 84 | 64 | 23 | 115 | 36.9 | 0.471 | 28 | 0 |
| 7 | 124 | 70 | 33 | 215 | 25.5 | 0.161 | 37 | 0 |
| 1 | 97 | 70 | 40 | 0 | 38.1 | 0.218 | 30 | 0 |
| 8 | 110 | 76 | 0 | 0 | 27.8 | 0.237 | 58 | 0 |
| 11 | 103 | 68 | 40 | 0 | 46.2 | 0.126 | 42 | 0 |
| 11 | 85 | 74 | 0 | 0 | 30.1 | 0.3 | 35 | 0 |
| 6 | 125 | 76 | 0 | 0 | 33.8 | 0.121 | 54 | 1 |
| 0 | 198 | 66 | 32 | 274 | 41.3 | 0.502 | 28 | 1 |
| 1 | 87 | 68 | 34 | 77 | 37.6 | 0.401 | 24 | 0 |
| 6 | 99 | 60 | 19 | 54 | 26.9 | 0.497 | 32 | 0 |
| 0 | 91 | 80 | 0 | 0 | 32.4 | 0.601 | 27 | 0 |
| 2 | 95 | 54 | 14 | 88 | 26.1 | 0.748 | 22 | 0 |
| 1 | 99 | 72 | 30 | 18 | 38.6 | 0.412 | 21 | 0 |
| 6 | 92 | 62 | 32 | 126 | 32 | 0.085 | 46 | 0 |
| 4 | 154 | 72 | 29 | 126 | 31.3 | 0.338 | 37 | 0 |
| 0 | 121 | 66 | 30 | 165 | 34.3 | 0.203 | 33 | 1 |
| 3 | 78 | 70 | 0 | 0 | 32.5 | 0.27 | 39 | 0 |
| 2 | 130 | 96 | 0 | 0 | 22.6 | 0.268 | 21 | 0 |
| 3 | 111 | 58 | 31 | 44 | 29.5 | 0.43 | 22 | 0 |
| 2 | 98 | 60 | 17 | 120 | 34.7 | 0.198 | 22 | 0 |
| 1 | 143 | 86 | 30 | 330 | 30.1 | 0.892 | 23 | 0 |
| 1 | 119 | 44 | 47 | 63 | 35.5 | 0.28 | 25 | 0 |
| 6 | 108 | 44 | 20 | 130 | 24 | 0.813 | 35 | 0 |
| 2 | 118 | 80 | 0 | 0 | 42.9 | 0.693 | 21 | 1 |
| 10 | 133 | 68 | 0 | 0 | 27 | 0.245 | 36 | 0 |
| 2 | 197 | 70 | 99 | 0 | 34.7 | 0.575 | 62 | 1 |
| 0 | 151 | 90 | 46 | 0 | 42.1 | 0.371 | 21 | 1 |
| 6 | 109 | 60 | 27 | 0 | 25 | 0.206 | 27 | 0 |
| 12 | 121 | 78 | 17 | 0 | 26.5 | 0.259 | 62 | 0 |
| 8 | 100 | 76 | 0 | 0 | 38.7 | 0.19 | 42 | 0 |
| 8 | 124 | 76 | 24 | 600 | 28.7 | 0.687 | 52 | 1 |
| 1 | 93 | 56 | 11 | 0 | 22.5 | 0.417 | 22 | 0 |
| 8 | 143 | 66 | 0 | 0 | 34.9 | 0.129 | 41 | 1 |
| 6 | 103 | 66 | 0 | 0 | 24.3 | 0.249 | 29 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 3 | 176 | 86 | 27 | 156 | 33.3 | 1.154 | 52 | 1 |
| 0 | 73 | 0 | 0 | 0 | 21.1 | 0.342 | 25 | 0 |
| 11 | 111 | 84 | 40 | 0 | 46.8 | 0.925 | 45 | 1 |
| 2 | 112 | 78 | 50 | 140 | 39.4 | 0.175 | 24 | 0 |
| 3 | 132 | 80 | 0 | 0 | 34.4 | 0.402 | 44 | 1 |
| 2 | 82 | 52 | 22 | 115 | 28.5 | 1.699 | 25 | 0 |
| 6 | 123 | 72 | 45 | 230 | 33.6 | 0.733 | 34 | 0 |
| 0 | 188 | 82 | 14 | 185 | 32 | 0.682 | 22 | 1 |
| 0 | 67 | 76 | 0 | 0 | 45.3 | 0.194 | 46 | 0 |
| 1 | 89 | 24 | 19 | 25 | 27.8 | 0.559 | 21 | 0 |
| 1 | 173 | 74 | 0 | 0 | 36.8 | 0.088 | 38 | 1 |
| 1 | 109 | 38 | 18 | 120 | 23.1 | 0.407 | 26 | 0 |
| 1 | 108 | 88 | 19 | 0 | 27.1 | 0.4 | 24 | 0 |
| 6 | 96 | 0 | 0 | 0 | 23.7 | 0.19 | 28 | 0 |
| 1 | 124 | 74 | 36 | 0 | 27.8 | 0.1 | 30 | 0 |
| 7 | 150 | 78 | 29 | 126 | 35.2 | 0.692 | 54 | 1 |
| 4 | 183 | 0 | 0 | 0 | 28.4 | 0.212 | 36 | 1 |
| 1 | 124 | 60 | 32 | 0 | 35.8 | 0.514 | 21 | 0 |
| 1 | 181 | 78 | 42 | 293 | 40 | 1.258 | 22 | 1 |
| 1 | 92 | 62 | 25 | 41 | 19.5 | 0.482 | 25 | 0 |
| 0 | 152 | 82 | 39 | 272 | 41.5 | 0.27 | 27 | 0 |
| 1 | 111 | 62 | 13 | 182 | 24 | 0.138 | 23 | 0 |
| 3 | 106 | 54 | 21 | 158 | 30.9 | 0.292 | 24 | 0 |
| 3 | 174 | 58 | 22 | 194 | 32.9 | 0.593 | 36 | 1 |
| 7 | 168 | 88 | 42 | 321 | 38.2 | 0.787 | 40 | 1 |
| 6 | 105 | 80 | 28 | 0 | 32.5 | 0.878 | 26 | 0 |
| 11 | 138 | 74 | 26 | 144 | 36.1 | 0.557 | 50 | 1 |
| 3 | 106 | 72 | 0 | 0 | 25.8 | 0.207 | 27 | 0 |
| 6 | 117 | 96 | 0 | 0 | 28.7 | 0.157 | 30 | 0 |
| 2 | 68 | 62 | 13 | 15 | 20.1 | 0.257 | 23 | 0 |
| 9 | 112 | 82 | 24 | 0 | 28.2 | 1.282 | 50 | 1 |
| 0 | 119 | 0 | 0 | 0 | 32.4 | 0.141 | 24 | 1 |
| 2 | 112 | 86 | 42 | 160 | 38.4 | 0.246 | 28 | 0 |
| 2 | 92 | 76 | 20 | 0 | 24.2 | 1.698 | 28 | 0 |
| 6 | 183 | 94 | 0 | 0 | 40.8 | 1.461 | 45 | 0 |
| 0 | 94 | 70 | 27 | 115 | 43.5 | 0.347 | 21 | 0 |
| 2 | 108 | 64 | 0 | 0 | 30.8 | 0.158 | 21 | 0 |
| 4 | 90 | 88 | 47 | 54 | 37.7 | 0.362 | 29 | 0 |
| 0 | 125 | 68 | 0 | 0 | 24.7 | 0.206 | 21 | 0 |
| 0 | 132 | 78 | 0 | 0 | 32.4 | 0.393 | 21 | 0 |
| 5 | 128 | 80 | 0 | 0 | 34.6 | 0.144 | 45 | 0 |
| 4 | 94 | 65 | 22 | 0 | 24.7 | 0.148 | 21 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 7 | 114 | 64 | 0 | 0 | 27.4 | 0.732 | 34 | 1 |
| 0 | 102 | 78 | 40 | 90 | 34.5 | 0.238 | 24 | 0 |
| 2 | 111 | 60 | 0 | 0 | 26.2 | 0.343 | 23 | 0 |
| 1 | 128 | 82 | 17 | 183 | 27.5 | 0.115 | 22 | 0 |
| 10 | 92 | 62 | 0 | 0 | 25.9 | 0.167 | 31 | 0 |
| 13 | 104 | 72 | 0 | 0 | 31.2 | 0.465 | 38 | 1 |
| 5 | 104 | 74 | 0 | 0 | 28.8 | 0.153 | 48 | 0 |
| 2 | 94 | 76 | 18 | 66 | 31.6 | 0.649 | 23 | 0 |
| 7 | 97 | 76 | 32 | 91 | 40.9 | 0.871 | 32 | 1 |
| 1 | 100 | 74 | 12 | 46 | 19.5 | 0.149 | 28 | 0 |
| 0 | 102 | 86 | 17 | 105 | 29.3 | 0.695 | 27 | 0 |
| 4 | 128 | 70 | 0 | 0 | 34.3 | 0.303 | 24 | 0 |
| 6 | 147 | 80 | 0 | 0 | 29.5 | 0.178 | 50 | 1 |
| 4 | 90 | 0 | 0 | 0 | 28 | 0.61 | 31 | 0 |
| 3 | 103 | 72 | 30 | 152 | 27.6 | 0.73 | 27 | 0 |
| 2 | 157 | 74 | 35 | 440 | 39.4 | 0.134 | 30 | 0 |
| 1 | 167 | 74 | 17 | 144 | 23.4 | 0.447 | 33 | 1 |
| 0 | 179 | 50 | 36 | 159 | 37.8 | 0.455 | 22 | 1 |
| 11 | 136 | 84 | 35 | 130 | 28.3 | 0.26 | 42 | 1 |
| 0 | 107 | 60 | 25 | 0 | 26.4 | 0.133 | 23 | 0 |
| 1 | 91 | 54 | 25 | 100 | 25.2 | 0.234 | 23 | 0 |
| 1 | 117 | 60 | 23 | 106 | 33.8 | 0.466 | 27 | 0 |
| 5 | 123 | 74 | 40 | 77 | 34.1 | 0.269 | 28 | 0 |
| 2 | 120 | 54 | 0 | 0 | 26.8 | 0.455 | 27 | 0 |
| 1 | 106 | 70 | 28 | 135 | 34.2 | 0.142 | 22 | 0 |
| 2 | 155 | 52 | 27 | 540 | 38.7 | 0.24 | 25 | 1 |
| 2 | 101 | 58 | 35 | 90 | 21.8 | 0.155 | 22 | 0 |
| 1 | 120 | 80 | 48 | 200 | 38.9 | 1.162 | 41 | 0 |
| 11 | 127 | 106 | 0 | 0 | 39 | 0.19 | 51 | 0 |
| 3 | 80 | 82 | 31 | 70 | 34.2 | 1.292 | 27 | 1 |
| 10 | 162 | 84 | 0 | 0 | 27.7 | 0.182 | 54 | 0 |
| 1 | 199 | 76 | 43 | 0 | 42.9 | 1.394 | 22 | 1 |
| 8 | 167 | 106 | 46 | 231 | 37.6 | 0.165 | 43 | 1 |
| 9 | 145 | 80 | 46 | 130 | 37.9 | 0.637 | 40 | 1 |
| 6 | 115 | 60 | 39 | 0 | 33.7 | 0.245 | 40 | 1 |
| 1 | 112 | 80 | 45 | 132 | 34.8 | 0.217 | 24 | 0 |
| 4 | 145 | 82 | 18 | 0 | 32.5 | 0.235 | 70 | 1 |
| 10 | 111 | 70 | 27 | 0 | 27.5 | 0.141 | 40 | 1 |
| 6 | 98 | 58 | 33 | 190 | 34 | 0.43 | 43 | 0 |
| 9 | 154 | 78 | 30 | 100 | 30.9 | 0.164 | 45 | 0 |
| 6 | 165 | 68 | 26 | 168 | 33.6 | 0.631 | 49 | 0 |
| 1 | 99 | 58 | 10 | 0 | 25.4 | 0.551 | 21 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 10 | 68 | 106 | 23 | 49 | 35.5 | 0.285 | 47 | 0 |
| 3 | 123 | 100 | 35 | 240 | 57.3 | 0.88 | 22 | 0 |
| 8 | 91 | 82 | 0 | 0 | 35.6 | 0.587 | 68 | 0 |
| 6 | 195 | 70 | 0 | 0 | 30.9 | 0.328 | 31 | 1 |
| 9 | 156 | 86 | 0 | 0 | 24.8 | 0.23 | 53 | 1 |
| 0 | 93 | 60 | 0 | 0 | 35.3 | 0.263 | 25 | 0 |
| 3 | 121 | 52 | 0 | 0 | 36 | 0.127 | 25 | 1 |
| 2 | 101 | 58 | 17 | 265 | 24.2 | 0.614 | 23 | 0 |
| 2 | 56 | 56 | 28 | 45 | 24.2 | 0.332 | 22 | 0 |
| 0 | 162 | 76 | 36 | 0 | 49.6 | 0.364 | 26 | 1 |
| 0 | 95 | 64 | 39 | 105 | 44.6 | 0.366 | 22 | 0 |
| 4 | 125 | 80 | 0 | 0 | 32.3 | 0.536 | 27 | 1 |
| 5 | 136 | 82 | 0 | 0 | 0 | 0.64 | 69 | 0 |
| 2 | 129 | 74 | 26 | 205 | 33.2 | 0.591 | 25 | 0 |
| 3 | 130 | 64 | 0 | 0 | 23.1 | 0.314 | 22 | 0 |
| 1 | 107 | 50 | 19 | 0 | 28.3 | 0.181 | 29 | 0 |
| 1 | 140 | 74 | 26 | 180 | 24.1 | 0.828 | 23 | 0 |
| 1 | 144 | 82 | 46 | 180 | 46.1 | 0.335 | 46 | 1 |
| 8 | 107 | 80 | 0 | 0 | 24.6 | 0.856 | 34 | 0 |
| 13 | 158 | 114 | 0 | 0 | 42.3 | 0.257 | 44 | 1 |
| 2 | 121 | 70 | 32 | 95 | 39.1 | 0.886 | 23 | 0 |
| 7 | 129 | 68 | 49 | 125 | 38.5 | 0.439 | 43 | 1 |
| 2 | 90 | 60 | 0 | 0 | 23.5 | 0.191 | 25 | 0 |
| 7 | 142 | 90 | 24 | 480 | 30.4 | 0.128 | 43 | 1 |
| 3 | 169 | 74 | 19 | 125 | 29.9 | 0.268 | 31 | 1 |
| 0 | 99 | 0 | 0 | 0 | 25 | 0.253 | 22 | 0 |
| 4 | 127 | 88 | 11 | 155 | 34.5 | 0.598 | 28 | 0 |
| 4 | 118 | 70 | 0 | 0 | 44.5 | 0.904 | 26 | 0 |
| 2 | 122 | 76 | 27 | 200 | 35.9 | 0.483 | 26 | 0 |
| 6 | 125 | 78 | 31 | 0 | 27.6 | 0.565 | 49 | 1 |
| 1 | 168 | 88 | 29 | 0 | 35 | 0.905 | 52 | 1 |
| 2 | 129 | 0 | 0 | 0 | 38.5 | 0.304 | 41 | 0 |
| 4 | 110 | 76 | 20 | 100 | 28.4 | 0.118 | 27 | 0 |
| 6 | 80 | 80 | 36 | 0 | 39.8 | 0.177 | 28 | 0 |
| 10 | 115 | 0 | 0 | 0 | 0 | 0.261 | 30 | 1 |
| 2 | 127 | 46 | 21 | 335 | 34.4 | 0.176 | 22 | 0 |
| 9 | 164 | 78 | 0 | 0 | 32.8 | 0.148 | 45 | 1 |
| 2 | 93 | 64 | 32 | 160 | 38 | 0.674 | 23 | 1 |
| 3 | 158 | 64 | 13 | 387 | 31.2 | 0.295 | 24 | 0 |
| 5 | 126 | 78 | 27 | 22 | 29.6 | 0.439 | 40 | 0 |
| 10 | 129 | 62 | 36 | 0 | 41.2 | 0.441 | 38 | 1 |
| 0 | 134 | 58 | 20 | 291 | 26.4 | 0.352 | 21 | 0 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 3 | 102 | 74 | 0 | 0 | 29.5 | 0.121 | 32 | 0 |
| 7 | 187 | 50 | 33 | 392 | 33.9 | 0.826 | 34 | 1 |
| 3 | 173 | 78 | 39 | 185 | 33.8 | 0.97 | 31 | 1 |
| 10 | 94 | 72 | 18 | 0 | 23.1 | 0.595 | 56 | 0 |
| 1 | 108 | 60 | 46 | 178 | 35.5 | 0.415 | 24 | 0 |
| 5 | 97 | 76 | 27 | 0 | 35.6 | 0.378 | 52 | 1 |
| 4 | 83 | 86 | 19 | 0 | 29.3 | 0.317 | 34 | 0 |
| 1 | 114 | 66 | 36 | 200 | 38.1 | 0.289 | 21 | 0 |
| 1 | 149 | 68 | 29 | 127 | 29.3 | 0.349 | 42 | 1 |
| 5 | 117 | 86 | 30 | 105 | 39.1 | 0.251 | 42 | 0 |
| 1 | 111 | 94 | 0 | 0 | 32.8 | 0.265 | 45 | 0 |
| 4 | 112 | 78 | 40 | 0 | 39.4 | 0.236 | 38 | 0 |
| 1 | 116 | 78 | 29 | 180 | 36.1 | 0.496 | 25 | 0 |
| 0 | 141 | 84 | 26 | 0 | 32.4 | 0.433 | 22 | 0 |
| 2 | 175 | 88 | 0 | 0 | 22.9 | 0.326 | 22 | 0 |
| 2 | 92 | 52 | 0 | 0 | 30.1 | 0.141 | 22 | 0 |
| 3 | 130 | 78 | 23 | 79 | 28.4 | 0.323 | 34 | 1 |
| 8 | 120 | 86 | 0 | 0 | 28.4 | 0.259 | 22 | 1 |
| 2 | 174 | 88 | 37 | 120 | 44.5 | 0.646 | 24 | 1 |
| 2 | 106 | 56 | 27 | 165 | 29 | 0.426 | 22 | 0 |
| 2 | 105 | 75 | 0 | 0 | 23.3 | 0.56 | 53 | 0 |
| 4 | 95 | 60 | 32 | 0 | 35.4 | 0.284 | 28 | 0 |
| 0 | 126 | 86 | 27 | 120 | 27.4 | 0.515 | 21 | 0 |
| 8 | 65 | 72 | 23 | 0 | 32 | 0.6 | 42 | 0 |
| 2 | 99 | 60 | 17 | 160 | 36.6 | 0.453 | 21 | 0 |
| 1 | 102 | 74 | 0 | 0 | 39.5 | 0.293 | 42 | 1 |
| 11 | 120 | 80 | 37 | 150 | 42.3 | 0.785 | 48 | 1 |
| 3 | 102 | 44 | 20 | 94 | 30.8 | 0.4 | 26 | 0 |
| 1 | 109 | 58 | 18 | 116 | 28.5 | 0.219 | 22 | 0 |
| 9 | 140 | 94 | 0 | 0 | 32.7 | 0.734 | 45 | 1 |
| 13 | 153 | 88 | 37 | 140 | 40.6 | 1.174 | 39 | 0 |
| 12 | 100 | 84 | 33 | 105 | 30 | 0.488 | 46 | 0 |
| 1 | 147 | 94 | 41 | 0 | 49.3 | 0.358 | 27 | 1 |
| 1 | 81 | 74 | 41 | 57 | 46.3 | 1.096 | 32 | 0 |
| 3 | 187 | 70 | 22 | 200 | 36.4 | 0.408 | 36 | 1 |
| 6 | 162 | 62 | 0 | 0 | 24.3 | 0.178 | 50 | 1 |
| 4 | 136 | 70 | 0 | 0 | 31.2 | 1.182 | 22 | 1 |
| 1 | 121 | 78 | 39 | 74 | 39 | 0.261 | 28 | 0 |
| 3 | 108 | 62 | 24 | 0 | 26 | 0.223 | 25 | 0 |
| 0 | 181 | 88 | 44 | 510 | 43.3 | 0.222 | 26 | 1 |
| 8 | 154 | 78 | 32 | 0 | 32.4 | 0.443 | 45 | 1 |
| 1 | 128 | 88 | 39 | 110 | 36.5 | 1.057 | 37 | 1 |

| Pregnancies | Glucose | Blood Pressure | Skin Thickness | Insulin | BMI | Diabetes Pedigree Function | Age | Outcome |
|-------------|---------|----------------|----------------|---------|------|----------------------------|-----|---------|
| 7 | 137 | 90 | 41 | 0 | 32 | 0.391 | 39 | 0 |
| 0 | 123 | 72 | 0 | 0 | 36.3 | 0.258 | 52 | 1 |
| 1 | 106 | 76 | 0 | 0 | 37.5 | 0.197 | 26 | 0 |
| 6 | 190 | 92 | 0 | 0 | 35.5 | 0.278 | 66 | 1 |
| 2 | 88 | 58 | 26 | 16 | 28.4 | 0.766 | 22 | 0 |
| 9 | 170 | 74 | 31 | 0 | 44 | 0.403 | 43 | 1 |
| 9 | 89 | 62 | 0 | 0 | 22.5 | 0.142 | 33 | 0 |
| 10 | 101 | 76 | 48 | 180 | 32.9 | 0.171 | 63 | 0 |
| 2 | 122 | 70 | 27 | 0 | 36.8 | 0.34 | 27 | 0 |
| 5 | 121 | 72 | 23 | 112 | 26.2 | 0.245 | 30 | 0 |
| 1 | 126 | 60 | 0 | 0 | 30.1 | 0.349 | 47 | 1 |
| 1 | 93 | 70 | 31 | 0 | 30.4 | 0.315 | 23 | 0 |