

BAB III

METODOLOGI PENELITIAN

3.1. Kedudukan dan Koordinasi

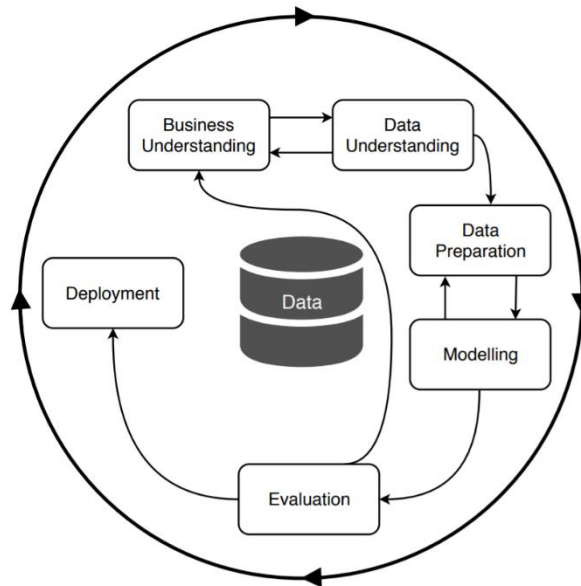
Pada pelaksanaan program kerja magang penelitian independen ini, dibutuhkan peran selayaknya *Data Scientist* antara lain sebagai berikut:

- a. Melakukan kegiatan *data mining* mulai dari pendefinisian masalah, pengumpulan data, pemilihan metode yang digunakan, pemrosesan data hingga menerapkan model klasifikasi *data mining* pada *Breast Coimbra Dataset* yang merupakan data hasil analisis darah dari salah satu rumah sakit di Portugal, yaitu *Coimbra Hospital and University Centre (CHUC)* untuk melakukan prediksi kanker payudara.
- b. Mengevaluasi model yang dibangun dan memilih model klasifikasi yang paling baik dari hasil penelitian.
- c. Melakukan penulisan jurnal ilmiah terkait penelitian dengan judul “Penerapan Algoritma *Data Mining Decision Tree* Untuk Prediksi Hasil Diagnosa Kanker Payudara”.

Pelaksanaan kerja magang ini dilakukan dengan bimbingan dan arahan dari bapak Iwan Prasetiawan sebagai dosen pembimbing lapangan yang mengawasi pelaksanaan program kerja magang ini.

3.2. Tugas yang Dilakukan

Pada program kerja magang penelitian independen ini, dilakukan beberapa kegiatan *data mining* dengan tahapan-tahapan yang ada pada metode CRISP-DM antara lain sebagai berikut:



Gambar 3.1 CRISP-DM

Sumber: [11]

1. *Business Understanding* adalah tahapan untuk memahami substansi dari aktivitas *data mining*, seperti menentukan tujuan atau capaian dari proyek *data mining* dan memahami situasi bisnis [12]. Pada tahap ini akan dijabarkan latar dari dilakukannya penelitian dan tujuan serta hal yang ingin dicapai dari kegiatan proyek *data mining* yang dilakukan melalui penelitian ini.
2. *Data Understanding* adalah tahapan untuk akuisisi data, mempelajari dan mendeskripsikan data, mengidentifikasi masalah terkait kualitas

data, dan mendeteksi bagian menarik dari data [12]. Pada tahap ini akan dijabarkan mengenai proses akuisisi data, penjelasan fitur-fitur pada data, pendeskripsian data dengan melakukan analisis deskriptif dan analisis korelasi fitur, dan mengidentifikasi kualitas data yang digunakan pada penelitian ini.

3. *Data Preparation* adalah tahapan untuk melakukan persiapan data meliputi pra-pemrosesan data, penyaringan data, dan ekstraksi fitur sesuai tujuan dari proyek *data mining* [13]. Pada tahapan ini akan dilakukan persiapan data dengan mengubah data menjadi bentuk format yang dapat digunakan untuk pembuatan model dan menentukan fitur-fitur yang akan digunakan pada pembuatan model.
4. *Modeling* adalah tahapan untuk menentukan teknik *data mining* meliputi menentukan algoritma, pencarian parameter yang sesuai algoritma yang digunakan, dan implementasi algoritma terhadap data untuk pelatihan model [13]. Pada tahapan ini akan dilakukan pembuatan model menggunakan beberapa algoritma klasifikasi data mining, yaitu *Decision Tree*, *Random Forest*, *K-Nearest Neighbor*, *Support Vector Machine*, dan *Naïve Bayes* serta dilakukan optimalisasi parameter untuk masing-masing algoritma.
5. *Evaluation* adalah tahapan untuk melakukan pengujian terhadap data baru [13]. Dari hasil pengujian tersebut akan dilakukan interpretasi dari hasil model data mining dan menilai hasil sudah sesuai dengan tujuan atau belum [12]. Pada tahapan ini, model-model klasifikasi yang telah

dibangun menggunakan masing-masing algoritma akan divalidasi dengan mengukur nilai akurasi dengan menggunakan teknik *5-fold cross validation*.

6. *Deployment* adalah tahapan untuk penerapan model yang sudah dibangun [13]. Pada tahap ini akan dipilih model klasifikasi terbaik untuk memprediksi kanker payudara pada kasus nyata.

3.3. Uraian Pelaksanaan Kerja Magang

Proses pengerjaan proyek dan pembangunan model dengan *framework* CRISP-DM dilakukan dalam durasi delapan minggu. Waktu pengerjaan dan kegiatan yang dilakukan secara detail dapat dilihat pada Tabel 3.1 berikut.

Tabel 3.1 Uraian Pelaksanaan Kerja Magang

Minggu Ke-	Kegiatan	Mulai	Selesai
1 dan 2	<i>Business Understanding</i>	21/09/2020	2/10/2020
2	<i>Data Understanding</i>	29/09/2020	2/10/2020
2	<i>Data Preprocessing</i>	1/10/2020	2/10/2020
3, 4, dan 5	<i>Data Modeling</i>	5/10/2020	23/10/2020
6 dan 7	<i>Evaluation</i>	26/10/2020	7/11/2020
8	<i>Deployment</i>	9/11/2020	13/11/2020

3.3.1. *Business Understanding*

Kanker Payudara merupakan jenis penyakit kanker kedua terbanyak yang menyerang wanita di seluruh dunia [1]. Tingkat kelangsungan hidup wanita penderita kanker payudara masih sangat rendah, khususnya di beberapa negara berkembang akibat kurangnya kemampuan prediksi sejak dini [2]. Hal ini berakibat fatal yang menyebabkan kanker payudara menjadi penyebab utama kematian akibat kanker pada wanita di seluruh dunia [3].

Untuk itu dibutuhkan kemampuan untuk memprediksi keberadaan kanker payudara pada tubuh sejak dini. Prediksi hasil diagnosis penyakit kanker payudara dapat menghasilkan prognosis yang baik dan tingkat kelangsungan hidup yang tinggi [14]. Salah satu bentuk pemeriksaan yang biasa dilakukan untuk mendiagnosis keberadaan kanker payudara, yaitu analisis darah, yang terdiri dari pemeriksaan *Age*, *Body Mass Index* (BMI), dan beberapa kandungan zat dalam tubuh, seperti *Glucose*, *Insulin*, HOMA, *Leptin*, *Adiponectin*, *Resistin*, dan MCP-1 [8]. Untuk dapat memberikan hasil analisis dan prediksi kanker payudara yang lebih optimal, data-data analisis darah yang dilakukan terhadap pasien-pasien sebelumnya dapat digunakan untuk membangun model *data mining* yang dapat memprediksi seseorang menderita kanker payudara atau tidak agar dapat segera dilakukan tindakan dan perawatan sedini mungkin untuk meningkatkan kelangsungan hidup penderita kanker payudara.

3.3.2. Data Understanding

Pada penelitian ini, data yang digunakan untuk membangun model prediksi kanker payudara adalah *Breast Cancer Coimbra Dataset* [8]. Data ini dapat diakses melalui internet pada <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra>. Data tersebut merupakan data hasil analisis darah pasien-pasien sebuah rumah sakit di Portugal, yaitu *Coimbra Hospital and University Centre* (CHUC). Data tersebut terdiri dari 10 atribut yang terdiri dari 9 atribut fitur dan 1

atribut target. Berikut atribut-atribut yang ada pada *Breast Cancer Coimbra Dataset* dapat dilihat pada Tabel.

Tabel 3.2 Atribut pada Dataset

Atribut	Keterangan
<i>Age</i>	Usia dari partisipan
<i>BMI</i>	Berat badan dari partisipan
<i>Glucose</i>	Kadar <i>glucose</i> (gula darah) pada tubuh partisipan
<i>Insulin</i>	Kadar <i>insulin</i> (hormon) pada tubuh partisipan
<i>HOMA</i>	Kadar HOMA pada tubuh partisipan
<i>Leptin</i>	Kadar <i>leptin</i> (hormon) pada tubuh partisipan
<i>Adiponectin</i>	Kadar <i>adiponectin</i> (hormon) pada tubuh partisipan
<i>Resistin</i>	Kadar <i>resistin</i> (<i>resistance to insulin</i>) pada tubuh partisipan
<i>MCP.1</i>	Kadar <i>MCP-1</i> (protein) pada tubuh partisipan
<i>Classification</i>	Kelas <i>Healthy</i> (Orang Sehat) atau <i>Unhealthy</i> (Penderita Kanker Payudara)

Berdasarkan fitur-fitur yang ada pada data tersebut, seseorang dapat diklasifikasikan menjadi dua kelas target “classification”, yaitu *healthy* (orang sehat) dan *unhealthy* (penderita kanker payudara). Pada dataset tersebut terdapat data hasil analisis darah dari 116 subjek yang terdiri dari 64 pasien penderita kanker payudara dan 52 orang yang sehat. Data pasien penderita kanker payudara dikumpulkan sebelum adanya perawatan terhadap penderita penyakit kanker payudara, sedangkan orang sehat dikumpulkan dari orang-orang yang tidak mengidap penyakit apapun [8].

Tabel 3.3 Contoh Sampel Data

	Age	BMI	Glucose	Insulin	HOMA	Leptin	Adiponectin	Resistin	MCP.1
0	48	23.50000	70	2.70700	0.46741	8.80710	9.70240	7.99585	417.11400
1	83	20.69049	92	3.11500	0.70690	8.84380	5.42929	4.06405	468.78600
2	82	23.12467	91	4.49800	1.00965	17.93930	22.43204	9.27715	554.69700
3	68	21.36752	77	3.22600	0.61272	9.88270	7.16956	12.76600	928.22000
4	86	21.11111	92	3.54900	0.80539	6.69940	4.81924	10.57635	773.92000
5	49	22.85446	92	3.22600	0.73209	6.83170	13.67975	10.31760	530.41000
6	89	22.70000	77	4.69000	0.89079	6.96400	5.58987	12.93610	1256.08300
7	76	23.80000	118	6.47000	1.88320	4.31100	13.25132	5.10420	280.69400
8	73	22.00000	97	3.35000	0.80154	4.47000	10.35872	6.28445	136.85500
9	75	23.00000	83	4.95200	1.01384	17.12700	11.57899	7.09130	318.30200
10	34	21.47000	78	3.46900	0.66744	14.57000	13.11000	6.92000	354.60000
11	29	23.01000	82	5.66300	1.14544	35.59000	26.72000	4.58000	174.80000
12	25	22.86000	82	4.09000	0.82727	20.45000	23.67000	5.14000	313.73000
13	24	18.67000	88	6.10700	1.33000	8.88000	36.06000	6.85000	632.22000
14	38	23.34000	75	5.78200	1.06967	15.26000	17.95000	9.35000	165.02000
15	44	20.76000	86	7.55300	1.60000	14.09000	20.32000	7.64000	63.61000
16	47	22.03000	84	2.86900	0.59000	26.65000	38.04000	3.32000	191.72000
17	61	32.03896	85	18.07700	3.79014	30.77290	7.78026	13.68392	444.39500

Analisis deskriptif dilakukan pada setiap fitur untuk melihat nilai sebaran pada fitur-fitur yang ada. Hasil analisis deskriptif tersebut dapat dilihat pada Tabel 3.4.

Tabel 3.4 Hasil Analisis Deskriptif Fitur

Fitur	Mean (Median)		
	Healthy	Unhealthy	All
<i>Age</i>	58 (65)	56,6 (53)	57,3 (56)
<i>BMI</i>	28,3 (27,7)	26,9 (27,4)	27,6 (27,6)
<i>Glucose</i>	88,2 (87)	105,6 (98,5)	97,8 (92)
<i>Insulin</i>	6,9 (5,4)	12,5 (7,6)	10 (5,9)
<i>HOMA</i>	1,5 (1,1)	3,6 (2)	2,7 (1,4)
<i>Leptin</i>	26,6 (21,5)	26,6 (18,9)	26,6 (20,3)
<i>Adiponectin</i>	10,3 (8,1)	10 (8,4)	10,2 (8,4)
<i>Resistin</i>	11,6 (8,9)	17,2 (14,4)	14,7 (10,8)
<i>MCP-1</i>	499,7 (471,3)	563 (465,4)	534,6 (471,3)

Berdasarkan Tabel 3.4 dapat dilihat terdapat beberapa atribut fitur yang membedakan antara orang yang sehat dengan penderita kanker payudara. Menurut perhitungan rata-rata dari data tersebut, ada lima fitur

yang secara signifikan berbeda antara orang yang sehat dengan penderita kanker payudara, yaitu kadar *Glucose* (19,7%), *Insulin* (81,2%), HOMA (140%), *Resistin* (49,1%), dan MCP-1 (12.67%) yang terdapat di dalam tubuh penderita kanker payudara lebih tinggi dibandingkan orang yang sehat.

Selanjutnya, analisis korelasi setiap fitur dengan kelas target juga dilakukan. Hasil analisis korelasi dapat dilihat pada Tabel 3.5.

Tabel 3.5 Hasil Analisis Korelasi Fitur

Fitur	Tingkat Korelasi
<i>Age</i>	-0.044
<i>BMI</i>	-0.133
<i>Glucose</i>	0.384
<i>Insulin</i>	0.277
<i>HOMA</i>	0.284
<i>Leptin</i>	-0.001
<i>Adiponectin</i>	-0.0195
<i>Resistin</i>	0.227
<i>MCP-1</i>	0.0914

Dari hasil analisis korelasi fitur pada Tabel 3.5 menunjukkan seberapa besar pengaruh antara setiap fitur yang ada dalam menentukan seseorang termasuk kelas orang sehat atau penderita kanker payudara. Dapat ditarik kesimpulan bahwa yang berpengaruh paling besar terhadap terjadinya kanker payudara adalah *Glucose* (0.384), HOMA (0.284), *Insulin* (0.277), dan *Resistin* (0.227). Pada Tabel 3.4 angka negatif menunjukkan bahwa hubungan yang terjadi antara fitur tersebut dengan kelas target bersifat negatif, artinya semakin tinggi nilai fitur maka semakin kecil kemungkinan terdapat kanker payudara, dan sebaliknya untuk angka positif.

3.3.3. Data Preparation

Pada tahapan ini, *data preprocessing* yang dilakukan adalah membagi data menjadi dua, yaitu *data training* dan *data testing* dengan rasio 80:20. Agar sesuai dengan penelitian sebelumnya [10] yang akan menjadi pembandingan tingkat akurasi dan karena jumlah data yang sedikit maka akan dilakukan pembagian data yang sama, yaitu *K-Fold Cross Validation* dengan $K=5$. Artinya data akan dibagi menjadi 5 bagian, 4 bagian untuk *training* (80%) dan 1 bagian untuk *testing* (20%) dan berulang hingga setiap bagian menjadi *data testing*. Semua fitur yang ada pada *dataset* digunakan dalam pembuatan model.

```
from sklearn.model_selection import KFold  
  
cv_5 = KFold(n_splits=5, shuffle=True, random_state=1000)
```

Gambar 3.2 Kode Python *K-Fold Cross Validation*

3.3.4. Modeling

Pada tahapan ini, pemodelan dilakukan menggunakan teknik klasifikasi untuk melakukan klasifikasi orang sehat dan penderita kanker payudara. Tahap pemodelan diawali dengan membandingkan beberapa algoritma klasifikasi *data mining* untuk memprediksi kanker payudara. Beberapa algoritma *data mining* yang dibandingkan adalah *Decision Tree* (DT), *Random Forest* (RF), *K-Nearest Neighbors* (k-NN), *Support Vector Machine* (SVM), dan *Naïve Bayes* (NB). Optimalisasi parameter juga

dilakukan untuk masing-masing algoritma yang digunakan untuk mencari model terbaik menggunakan teknik *Grid Search*.

```
def dtree(X, y):
    parameters = {'criterion': ['gini'],
                  'max_depth': [int(x) for x in np.linspace(start=2, stop=24, num=23)],
                  'min_samples_split': [int(x) for x in np.linspace(start=2, stop=40, num=39)],
                  'min_samples_leaf': [int(x) for x in np.linspace(start=1, stop=21, num=21)],
                  'max_leaf_nodes': [int(x) for x in np.linspace(start=2, stop=50, num=25)]
    }
    model = DecisionTreeClassifier()
    cv_5 = KFold(n_splits=5, shuffle=True, random_state=1000)
    gs = GridSearchCV(estimator=model, param_grid=parameters, cv=cv_5, scoring='accuracy', refit=False, verbose=10)
    gs.fit(X, y)
    return gs

def rforest(X, y):
    parameters = {'n_estimators': [int(x) for x in np.linspace(start=500, stop=2000, num=10)],
                  'max_features': ['auto', 'sqrt'],
                  'max_depth': [int(x) for x in np.linspace(10, 110, num=10)],
                  'min_samples_split': [2, 5, 10],
                  'min_samples_leaf': [1, 2, 4, 5, 10],
                  'bootstrap': [True, False]}
    model = RandomForestClassifier()
    cv_5 = KFold(n_splits=5, shuffle=True, random_state=1000)
    gs = GridSearchCV(estimator=model, param_grid=parameters, cv=cv_5, scoring='accuracy', refit=False, verbose=10)
    gs.fit(X, y)
    return gs

def svm(X, y):
    parameters = {'C': [0.1, 1, 10, 100, 1000],
                  'gamma': [1, 0.1, 0.01, 0.001, 0.0001],
                  'kernel': ['linear', 'rbf']}
    model = SVC()
    cv_5 = KFold(n_splits=5, shuffle=True, random_state=1000)
    gs = GridSearchCV(estimator=model, param_grid=parameters, cv=cv_5, scoring='accuracy', refit=False, verbose=10)
    gs.fit(X, y)
    return gs

def knn(X, y):
    parameters = {'n_neighbors': [3,5,11,19,21,23,25,27,29,31,33,35,37],
                  'weights': ['uniform', 'distance'],
                  'metric': ['euclidean', 'manhattan']}
    model = KNeighborsClassifier()
    cv_5 = KFold(n_splits=5, shuffle=True, random_state=1000)
    gs = GridSearchCV(estimator=model, param_grid=parameters, cv=cv_5, scoring='accuracy', refit=False, verbose=10)
    gs.fit(X, y)
    return gs

def nbayes(X, y):
    parameters = {'var_smoothing': np.logspace(0, -9, num=100)}
    model = GaussianNB()
    cv_5 = KFold(n_splits=5, shuffle=True, random_state=1000)
    gs = GridSearchCV(estimator=model, param_grid=parameters, cv=cv_5, scoring='accuracy', verbose=10)
    gs.fit(X, y)
    return gs
```

Gambar 3.3 Kode Python Optimalisasi Parameter

Hasil dari optimalisasi parameter dapat dilihat pada Tabel 3.6.

Tabel 3.6 Hasil Optimalisasi Parameter

Algoritma	Parameter	
	Nama Parameter	Nilai Terbaik
<i>Decision Tree</i>	<i>Criterion</i>	<i>Gini</i>
	<i>Max Depth</i>	4
	<i>Min Samples Split</i>	2
	<i>Min Samples Leaf</i>	1
	<i>Max Leaf Nodes</i>	10
<i>Random Forest</i>	<i>N estimators</i>	500
	<i>Max features</i>	<i>Auto</i>
	<i>Max depth</i>	87
	<i>Min samples split</i>	5
	<i>Min samples leaf</i>	1
	<i>Bootstrap</i>	<i>False</i>
<i>KNN</i>	<i>N neighbors</i>	1
	<i>Weights</i>	<i>Uniform</i>
	<i>Metric</i>	<i>Euclidean</i>
<i>SVM</i>	<i>C</i>	1
	<i>Gamma</i>	1
	<i>Kernel</i>	<i>Linear</i>
<i>Naïve Bayes</i>	<i>Var smoothing</i>	5.3366992312063

Dari hasil optimalisasi parameter, model dibangun menggunakan algoritma-algoritma klasifikasi dengan parameter terbaik yang telah didapatkan. Selanjutnya, model terbaik dari masing-masing algoritma akan dievaluasi dan dibandingkan pada tahapan berikutnya.

3.3.5. Evaluation

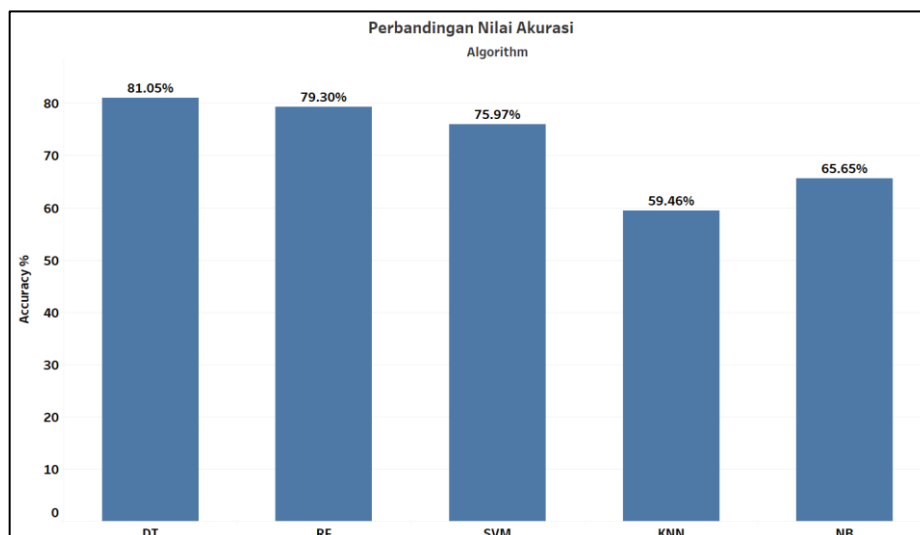
Pada tahapan ini, model-model yang sudah dilatih menggunakan *data training* dan parameter terbaik dievaluasi menggunakan *data testing*. Evaluasi model dilakukan menggunakan teknik *5-Fold Cross Validation*.

Hasil akurasi terbaik dari masing-masing model dapat dilihat pada Tabel 3.7 berikut ini.

Tabel 3.7 Nilai Akurasi Model

Algoritma	Tingkat Akurasi
<i>Decision Tree</i>	81,05%
<i>Random Forest</i>	79,30%
<i>KNN</i>	59,46%
<i>SVM</i>	75,97%
<i>Naïve Bayes</i>	65,65%

Dari hasil akurasi masing-masing model yang dibangun dapat disimpulkan bahwa *Decision Tree* merupakan algoritma yang memiliki performa terbaik dalam menentukan seseorang termasuk kelas orang sehat atau penderita kanker payudara berdasarkan hasil analisis darah. *Random Forest* juga memiliki performa yang hampir sama dengan *Decision Tree*, sedangkan SVM, KNN, dan *Naïve Bayes* memiliki performa yang kurang baik dibandingkan *Decision Tree*.



Gambar 3.4 Visualisasi Perbandingan Hasil Nilai Akurasi

Hasil akurasi dari masing-masing model yang dibangun pada penelitian ini juga dibandingkan dengan penelitian-penelitian yang sudah ada sebelumnya. Perbandingan dengan penelitian sebelumnya dapat dilihat pada Tabel 3.8.

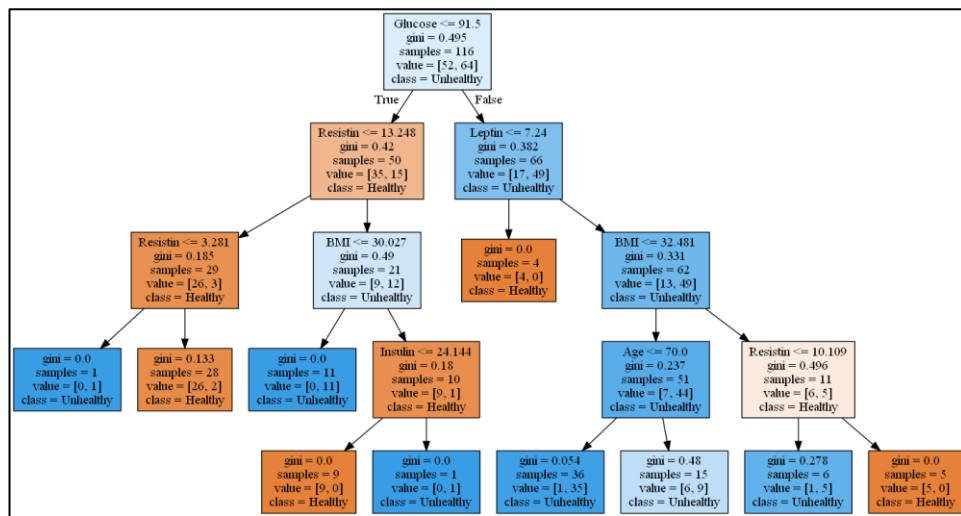
Tabel 3.8 Perbandingan Nilai Akurasi dengan Penelitian Lain

Algoritma	Penelitian Lain	Penelitian Ini
<i>Decision Tree</i>	69% [10]	81,05%
	70% [15]	
	72% [9]	
<i>Random Forest</i>	66% [9]	79,30%
	70% [15]	
	70% [10]	
	74% [16]	
<i>KNN</i>	48% [15]	59,46%
	58% [10]	
<i>SVM</i>	71% [16]	75,97%
	72% [10]	
	73% [17]	
<i>Naïve Bayes</i>	62% [10]	65,65%
	66% [15]	
	68% [9]	

Dari hasil perbandingan performa akurasi model-model yang dibangun pada penelitian ini dengan penelitian sebelumnya, dapat disimpulkan bahwa penelitian ini telah membangun model yang memiliki performa yang lebih baik. Model *Decision Tree* yang dibangun pada penelitian ini memiliki performa yang lebih baik dibandingkan model dengan algoritma yang lain. Untuk itu, model *Decision Tree* ini dapat digunakan pada tahap *deployment* untuk memprediksi data baru analisis darah.

3.3.6. Deployment

Setelah melakukan evaluasi terhadap model-model yang dibangun, didapatkan model terbaik, yaitu *Decision Tree* dengan akurasi 81,05%. Model *Decision Tree* yang dibangun dapat dilihat pada Gambar 3.5 berikut ini.



Gambar 3.5 Model Decision Tree

Dapat dilihat dari hasil prediksi menggunakan model algoritma *Decision Tree* pada Gambar 4.1, setiap *node* selalu memiliki empat komponen, yaitu *feature*, *samples*, *value*, dan *class*. *Feature* merupakan variabel yang menjadi indikasi seseorang *unhealthy* (penderita) atau *healthy* (sehat). *Samples* adalah jumlah data yang ada. *Value* menunjukkan jumlah masing-masing kategori pada *node* tersebut. *Class* merupakan label yang menunjukkan seseorang berada pada kelas *unhealthy* atau *healthy*.

Melalui Gambar 4.1, beberapa informasi yang dapat diketahui dari *root node*, yaitu pada *stage* awal model terdapat 116 jumlah *samples*,

dengan pembagian 52 *healthy* dan 64 *unhealthy*. Variabel yang paling berpengaruh terhadap diagnosa seseorang terkena kanker payudara atau tidak adalah *Glucose*. Hal ini ditunjukkan dengan pemilihan *Glucose* sebagai *rule* awal pada *root node* dengan nilai *gini index* terbesar di antara variabel lainnya, yaitu 0.495. Beberapa fitur penentu seseorang terserang penyakit kanker payudara atau tidak yang disajikan melalui Tabel 3.9 sebagai berikut.

Tabel 3. 9 Fitur Penentu Berdasarkan Gini Index

Node	Fitur	Gini Index
<i>Root Node (Level 0)</i>	Glucose	0.495
<i>Node 1 (Level 1)</i>	Resistin	0.42
<i>Node 2 (Level 1)</i>	Leptin	0.382
<i>Node 1 (Level 2)</i>	Resistin	0.185
<i>Node 2 (Level 2)</i>	BMI	0.49
<i>Node 3 (Level 2)</i>	BMI	0.331
<i>Node 1 (Level 3)</i>	Insulin	0.18
<i>Node 2 (Level 3)</i>	Age	0.237
<i>Node 3 (Level 3)</i>	Resistin	0.496

3.4. Kendala yang Dihadapi

Beberapa kendala yang dihadapi dalam pelaksanaan kerja magang antara lain sebagai berikut:

1. Adanya kesulitan untuk mendapatkan sumber data atau mengumpulkan data mengenai kanker payudara dalam kurun waktu lima tahun terakhir.
2. Proses *training* model membutuhkan sumber daya RAM yang cukup besar dan waktu yang cukup lama.
3. Program kerja magang proyek independen ini merupakan *Batch 1*, sehingga informasi yang tersedia tentang program tersebut masih

sedikit. Selain itu, terdapat banyak perubahan dari segi *timeline* karena disesuaikan dengan peserta proyek independen yang lain.

3.5. Solusi atas Kendala

Beberapa solusi yang dapat dilakukan dalam mengatasi kendala di atas antara lain sebagai berikut:

1. Menggunakan salah satu data yang ada dalam kurun waktu lima tahun terakhir tentang kanker payudara dan data yang digunakan pada penelitian ini adalah *Breast Cancer Coimbra Dataset* [8].
2. *Training* model yang sedang berjalan akan menyebabkan segala proses pada perangkat menjadi lambat karena konsumsi RAM mencapai 80% dari total 8 GB dan waktu yang dibutuhkan bisa mencapai 6-7 jam. Oleh karena itu, proses *training* model dilakukan pada malam hari ketika tidak ada aktivitas yang dilakukan pada perangkat.
3. Lebih aktif bertanya kepada pihak *Student Development dan Human Resources Development* Universitas Multimedia Nusantara agar memperoleh informasi mengenai prosedur pelaksanaan kerja magang.