

BAB II

LANDASAN TEORI

2.1 Tinjauan Teori

2.1.1 Penyakit Kardiovaskular dan Penyakit Jantung Koroner

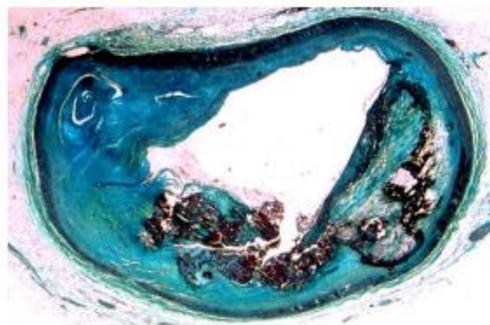
Penyakit kardiovaskular atau yang biasa disebut dengan *CVD* merupakan suatu keadaan dimana terjadinya gangguan pada sistem jantung dan pembuluh darah[17]. Dengan total kematian mencapai 13 juta kematian pada tahun 2010, Penyakit kardiovaskular telah menjadi penyebab utama kematian secara internasional[18]. Menurut WHO, korban kematian yang diakibatkan dari penyakit kardiovaskular berasal dari negara dengan pendapatan menengah hingga rendah atau bisa dikatakan sebagai negara berkembang[19]. Terdapat beberapa jenis penyakit kardiovaskular, dimana seluruh jenisnya berhubungan dengan gangguan jantung dan pembuluh darah seperti penyakit jantung koroner, *cerebrovascular*, *peripheral arterial*, *rheumatic heart disease*, *congenital heart disease*, *deep vein thrombosis* dan *pulmonary embolism*[4]. Penyakit kardiovaskular dapat dicegah dengan secara rutin mengontrol faktor utama penyebab penyakit kardiovaskular diantaranya adalah menjaga berat badan ideal, kolesterol dan lemak darah tetap pada tingkat normal dan menghindari penggunaan tembakau / rokok serta tidak mengkonsumsi alkohol[20].

Diantara jenis penyakit kardiovaskular, penyakit jantung koroner merupakan salah satu penyakit jantung yang cukup berbahaya karena berada pada sistem sistemik[21]. Penyakit jantung koroner sendiri adalah penyakit yang terjadi akibat suplai darah jantung tersumbat atau terganggu oleh penumpukan zat lemak di arteri koroner[22]. Penumpukan lemak yang terjadi pada arteri ini biasa disebut dengan plak (*plaque*)[23]. Penumpukan plak akan menyebabkan bagian dalam arteri koroner menyempit dari waktu ke waktu, yang menyebabkan aliran darah

terhalang, proses ini disebut aterosklerosis[24]. Penyakit jantung koroner tidak hanya terjadi pada usia lanjut saja. Penyakit ini juga menyerang mereka yang usianya lebih muda[25]. Berdasarkan hasil riset “Kesehatan Dasar” dari Kementerian Kesehatan Indonesia pada tahun 2013, di Indonesia sendiri tercatat 2.650.340 (1,5%) penderita penyakit jantung koroner dengan Provinsi Jawa Timur menjadi daerah dengan kasus penyakit jantung koroner tertinggi sebesar 375.127 (1,3%) pasien[26].

Gejala awal yang dapat dirasakan oleh pengidap penyakit jantung koroner dan pertanda bahwa penyakit telah memberikan efek yang cukup buruk pada jantung adalah *Angina* atau yang biasa dikenal sebagai nyeri dada (*chest pain*), sakit kepala, nyeri punggung dan juga sesak nafas[24]. Kondisi terburuk dari penyakit jantung koroner disebut dengan sindroma koroner akut (SKA).

Plak yang pecah

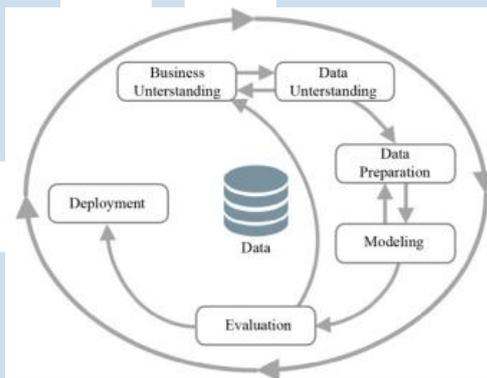


Gambar 2.1 Plak pada arteri jantung

SKA merupakan serangan jantung yang diakibatkan pecahnya plak aterosklerosis seperti yang dapat dilihat pada **gambar 2.1**. Hal ini dapat menyebabkan terbentuknya bekuan darah didalam pembuluh darah (arteri koroner) yang akan menyebabkan kemungkinan terburuk pada pasien berupa tidak berfungsinya otot jantung dan kematian[27]. Melihat bahaya yang dapat disebabkan dari penyakit jantung koroner, maka dari itu, sangat penting untuk menjaga gaya hidup sehari – hari sejak dini agar kesehatan jantung tetap terjaga[28].

2.1.2 CRISP – DM

Cross Industry Standard Process – Data Mining (CRISP - DM) adalah sebuah kerangka kerja yang memiliki fungsi sebagai penerjemah masalah seputar bisnis ke dalam data *mining*[29]. CRISP – DM merupakan salah satu metode data *mining process* yang banyak digunakan oleh mayoritas organisasi / perusahaan. Hal ini dikarenakan CRISP – DM memiliki prosedur yang bertahap dalam penerapannya sehingga *framework* yang dibangun menjadi lebih terstruktur[30].



Gambar 2.2 *Cross Industry Standard Process – Data Mining*

Seperti yang ada pada **gambar 2.2** diatas, dalam penerapannya, CRISP – DM terbagi menjadi beberapa langkah diantaranya adalah *Business understanding, Data understanding, Data preparation, Modelling, Evaluation, and Deployment*[31]. Adapun penjelasan lengkap mengenai ke enam langkah CRISP – DM ini adalah sebagai berikut [30]:

a) *Business understanding*

Tahapan ini merupakan tahapan yang mewakili langkah inti untuk melakukan *task Data Mining* yang baik. Dimana, *task* akan dianalisis, dipahami dan didefinisikan. Pada tahapan ini juga, yang akan menentukan masalah yang terjadi dan tujuan dari *Data Mining* yang dilakukan.

b) *Data understanding*

Tahapan *data understanding* dicirikan dengan dilakukannya pengumpulan data mentah yang akan digunakan dalam proses *Data Mining* dan juga merupakan pengenalan dengan data

mentah secara lebih detail. Pengenalan yang dilakukan termasuk dalam menganalisa kualitas *variable* yang terdapat pada data mentah. Analisa kualitas data dilakukan untuk menghilangkan *variable* yang tidak berhubungan dengan parameter yang digunakan pada proses *Data Mining*, atau *variable* yang tidak diketahui dalam kumpulan data besar, serta menghilangkan kumpulan data yang tidak diperlukan.

c) ***Data preparation***

Setelah data telah dipahami dengan baik, maka data mentah akan disiapkan sesuai dengan parameter dan tujuan dari *Data Mining*. Persiapan data yang dimaksudkan termasuk kedalam fase analisis siklus DMAIC (*Define, Measure, Analyze, Improve, Control*), yang dimulai dari seleksi *variable* dan parameter, pembersihan data, konstruksi, integrasi, dan pemformatan data sebagai persiapan untuk tahap pemodelan.

d) ***Modelling***

Tahapan ini merupakan bagian dari tahapan analisis yang berguna untuk tahapan pengujian perencanaan. Dan tahapan pemodelan dalam proyek *Data Mining* jauh lebih kompleks dibandingkan dengan tahapan – tahapan lainnya karena sebagian besar implementasi dari tahapan ini adalah menggunakan teknik pemodelan dan algoritma – algoritma *Data Mining*.

e) ***Evaluation***

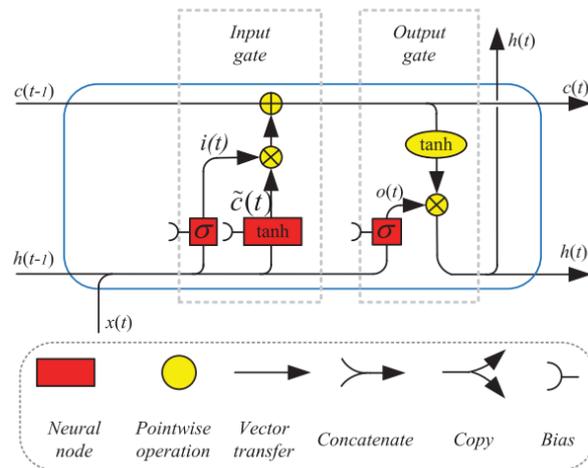
Pada tahapan ini evaluasi yang dimaksudkan adalah usaha yang dilakukan untuk meningkatkan seluruh proses dalam DMAIC. Hasil dari tahapan ini akan dievaluasi sesuai dengan tujuan dilakukannya *Data Mining*, model yang digunakan untuk menganalisa, dan manfaat dari analisa yang dilakukan.

f) **Deployment**

Tahapan ini berfokus kepada bagian perbaikan dari proses Data Mining yang telah dilakukan. Serta meliputi rencana proses dan loop kontrol yang dapat didefinisikan dalam proses perbaikan dan proses dokumentasi mengenai proyek Data Mining lain dimasa depan. Hal ini berarti bahwa pengetahuan yang diperoleh dari hasil analisa pada proses Data Mining akan disimpan kedalam sistem manajemen pengetahuan organisasi.

2.1.3 **Single Long Short Term Memory (LSTM)**

Seperti yang dapat dilihat pada gambar 2.3, algoritma Long Short Term Memory merupakan suatu algoritma jenis khusus yang masih merupakan bagian dari RNN (Recurrent Neural Network) yang sangat kompeten dalam learning long-term[32].



Gambar 2.3 Framework LSTM

Long Short Term Memory memungkinkan network untuk mempertahankan apa yang dipelajarinya dalam jangka waktu yang lebih lama. Hal ini menjadikan algoritma Long Short Term Memory telah terbukti memiliki kinerja yang sangat baik pada tugas seperti speech recognition, acoustic modeling, sentence embedding dan lainnya[33]. Dalam menggunakan algoritma Long Short Term Memory untuk melakukan training model, panjang dari dataset perlu dipertimbangkan.

Untuk memanfaatkan karakteristik temporal *Long Short Term Memory*, p_i harus diperpanjang dengan panjang yang sama dengan X_i , $P_i = [p_i, p_i, \dots, p_i]$, di mana baris panjang vektor yang dimaksudkan adalah l [34]. Singkatnya, seluruh perhitungan algoritma *Long Short Term Memory* dapat digambarkan oleh deret persamaan pada **rumus 2.1** berikut ini [35]:

$$\begin{aligned}
 f_{(t)} &= \sigma(W_{fx}x_{(t)} + W_{fh}h_{(t-1)} + b_f) \\
 i_{(t)} &= \sigma(W_{ix}x_{(t)} + W_{ih}h_{(t-1)} + b_i) \\
 o_{(t)} &= \sigma(W_{ox}x_{(t)} + W_{oh}h_{(t-1)} + b_o) \\
 c_{(t)} &= \tanh(W_{cx}x_{(t)} + W_{ch}h_{(t-1)} + b_c)
 \end{aligned}$$

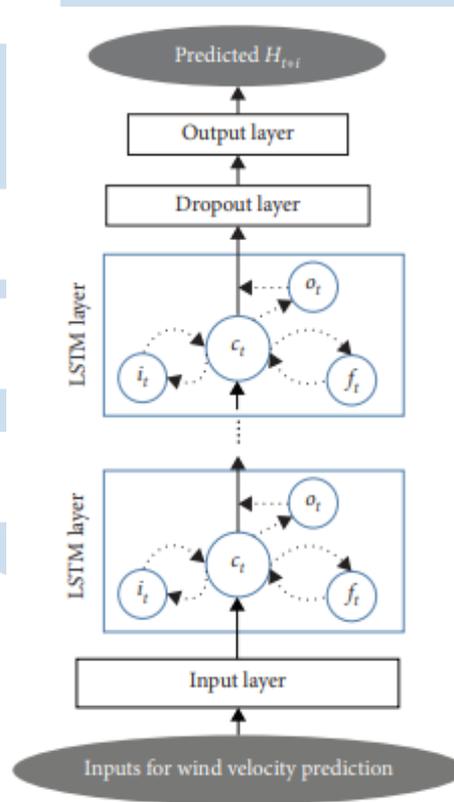
Rumus 2.1 Persamaan LSTM

Dimana W_x dan W_h adalah bobot dan b adalah vektor bias. Fungsi aktivasi ditunjuk oleh σ yang merupakan sigmoidnya dan berfungsi sebagai gerbang umum yang dapat digunakan sebagai aktivasi fungsi. Dan *Tanh* mewakili fungsi aktivasi nonlinier[35].

2.1.4 Stacked Long Short Term Memory (LSTM)

Berbeda dengan Algoritma *Long Short Term Memory*, *Stacked Long Short Term Memory* merupakan algoritma yang menerapkan arsitektur *Long Short Term Memory* secara bertumpuk atau yang biasa didefinisikan sebagai *Long Short Term Memory* dengan model yang terdiri dari beberapa lapisan *Long Short Term Memory* seperti yang dapat dilihat pada **gambar 2.4**. *Stacked Long Short Term Memory* pertama kali dirumuskan dan diterapkan untuk keperluan *speech recognition* (pengenalan suara). Model *Stacked Long Short Term Memory* menggunakan beberapa lapisan model *Long Short Term Memory* yang ditumpuk sebelum diteruskan kepada *output* akhir. Dalam *Stacked Long Short Term Memory* sendiri, lapisan model *Long Short Term Memory* pertama akan menghasilkan vektor urutan yang digunakan sebagai *input*

dari lapisan model *Long Short Term Memory* berikutnya. Selain itu, *Stacked Long Short Term Memory* menerima umpan balik dari model *Long Short Term Memory* sebelumnya, yang memungkinkan model untuk menangkap pola pada *dataset* yang digunakan. Hal ini juga berguna untuk menghindari kemungkinan *overfitting* pada *dataset*[36].



Gambar 2.4 Proses Stacked-LSTM

2.1.5 Confusion Matrix

Confusion Matrix merupakan suatu metode evaluasi yang biasa digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Dengan menggunakan *confusion matrix* maka akan diketahui tingkat nilai akurasi, presisi dan *recall* pada model yang telah dibangun[37]. Struktur kuadrat dari *confusion matrix* direpresentasikan melalui baris dan kolom, di mana baris adalah kelas aktual dari *instance*, dan kolom adalah kelas data yang dianalisa[38].

2.1.6 Python

Bahasa Pemrograman *Python* adalah salah satu bahasa pemrograman yang dapat dikatakan mirip seperti MATLAB. Namun, lebih daripada MATLAB, Bahasa pemrograman python juga memiliki fitur *object oriented programming* (OOP) dan juga menyediakan antarmuka lintas *platform*[39]. Kemampuan *python* diperluas melalui kumpulan *package* yang ada pada *library python* yang memiliki lebih dari 100.000 *package* dengan fungsi yang berbeda – beda[40]. Dimana dengan *package - package* yang telah disediakan oleh *python* tersebut, para *developer* dapat memperoleh lebih banyak fleksibilitas dan kemudahan untuk menulis model yang dibangun[39].

2.1.7 Google Colaboratory

Google Colaboratory atau yang biasa disebut dengan *Google Colab* merupakan sebuah program *executable document* yang memiliki banyak kegunaan diantaranya adalah untuk menulis, menjalankan dan menyimpan *syntax* dari bahasa pemrograman *python* dimana dokumen *python* yang telah ditulis dapat membagikan oleh kesesama pengguna akun *google* melalui *Google Drive*[41].

2.2 Penelitian Terdahulu

Adapun penelitian sebelumnya terkait dengan analisis penyakit jantung koroner yang dijadikan acuan dilakukannya penelitian ini dapat dilihat pada tabel 2.1 dibawah ini.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Tabel 2.1 Penelitian terdahulu

Nama Jurnal	Nama, Artikel (Vol,Tahun)	Nama Penulis; Tahun	Masalah	Metode	Hasil
2nd International Conference on Information Technology and Computer Application (ITCA) , 2020	“Heart disease prediction based on random forest and LSTM” (2020)[15]	(Yuepeng Liu, Mengfei Zhang, Zezhong Fan, Yinghan Chen) ;2020	Penyakit jantung telah menjadi ancaman utama bagi kesehatan orang-orang di seluruh dunia. Sementara sebagian besar penyakit jantung seharusnya dapat dicegah dan dikendalikan secara efektif.	Pencegahan penyakit jantung dapat dilakukan dengan Memanfaatkan perkembangan teknologi komputer dan <i>machine learning</i> yang dapat membantu para tenaga medis untuk membuat keputusan diagnosis terhadap pasien.	Prediksi dengan menggunakan algoritma LSTM menghasilkan akurasi sebesar 84.5%

Nama Jurnal	Nama, Artikel (Vol, Tahun)	Nama Penulis; Tahun	Masalah	Metode	Hasil
Ultimatics : Jurnal Teknik Informatika, 2020	“Prediksi Harga Beras Menggunakan Metode Multilayer Perceptron (MLP) dan Long Short Term Memory (LSTM)” (12, 2020)[42]	(Steven Sen, Dedy Sugiarto, Abdul Rochman;2020)	Pada masa sulit di era Covid19 ini harga-harga pangan pastinya diharapkan tetap stabil demi menunjang kehidupan masyarakat, terutama beras yang merupakan keperluan pokok warna negara Indonesia. Sehingga diperlukan adanya Prediksi terhadap harga beras yang berguna untuk mewujudkan kestabilan harga bahan pokok.	Prediksi akan harga beras di Indonesia ini dapat dilakukan dengan menggunakan beberapa algoritma <i>Machine Learning</i> seperti <i>Multilayer Perceptron</i> (MLP), dan <i>Long Short Term Memory</i> (LSTM)	Dapat disimpulkan bahwa algoritma LSTM akan lebih akurat jika dilihat dari rentang dengan harga aktual serta nilai RMSE- yang dihasilkan LSTM lebih kecil dibandingkan MLP.

Nama Jurnal	Nama, Artikel (Vol, Tahun)	Nama Penulis; Tahun	Masalah	Metode	Hasil
Health Information Science and Systems	"A stacked LSTM for atrial fibrillation prediction based on multivariate ECGs"(8, 2020)[43]	(Le Sun, Yukang Wang, Jinyuan He, Haoyuan Li, Dandan Peng, Yilin Wang; 2020)	Fibrilasi atrium (AF) adalah detak jantung yang tidak teratur dan cepat yang dapat meningkatkan risiko berbagai komplikasi terkait jantung, seperti stroke dan gagal jantung. Elektrokardiografi (EKG) banyak digunakan untuk memantau kesehatan pasien penyakit jantung. Ini dapat secara dramatis meningkatkan kesehatan dan tingkat	Pada penelitian ini, akan digunakan Stacked RNN yang terdiri dari LSTM bertumpuk untuk memprediksi AF, yang disebut SLAP. Model ini dapat secara efektif menghindari ledakan gradien dan ledakan gradien	Dari penelitian yang dilakukan eksperimen komprehensif berdasarkan dua kumpulan data publik. Hasil eksperimen ini menunjukkan akurasi 92% dan f-score 92% dari prediksi AF, yang lebih baik

Nama Jurnal	Nama, Artikel (Vol, Tahun)	Nama Penulis; Tahun	Masalah	Metode	Hasil
			kelangsungan hidup pasien penyakit jantung dengan secara akurat memprediksi AF dalam EKG	RNN biasa dan mempelajari fitur-fiturnya dengan lebih baik	daripada arsitektur deteksi AF lainnya seperti RNN dan LSTM
Conference: The 8th International Conference on Engineering, Technology, and Industrial Application	“Prediction Models for Coronary Heart Disease” (2021) [44]	(Cristiana Neto, Diana Ferreira, Jos´e Ramos, Sandro Cruz, Joaquim Oliveira, Ant´onio Abelha, Jos´e Machado ; 2021)	Penggunaan machine learning dapat menjadi dasar dalam memprediksi resiko terhadap penyakit jantung yang efektif dan akurat. Dimana penyakit jantung merupakan penyebab kematian terbesar didunia.	Pada penelitian ini akan membandingkan beberapa algoritma machine learning, diantaranya Logistic regression, K-nearest neighbor, ANN, SVM, Naive bayes,	CRISP DM memegang peranan penting dalam bidang kesehatan, karena dapat hasil analisa yang digunakan dapat meningkat

Nama Jurnal	Nama, Artikel (Vol, Tahun)	Nama Penulis; Tahun	Masalah	Metode	Hasil
				<p><i>Decision tree</i> dan juga <i>random forest</i></p>	<p>tkan kualitas hidup pasien dan juga dapat menyelamatkan nyawa. Studi ini menunjukkan peran CRISP-DM dalam prediksi risiko penyakit jantung koroner dari faktort berbeda - beda</p>

Nama Jurnal	Nama, Artikel (Vol, Tahun)	Nama Penulis; Tahun	Masalah	Metode	Hasil
Ultimatics : Jurnal Teknik Informatika	“Prediksi Tingkat Kelulusan Tepat Waktu Mahasiswa Menggunakan Algoritma Naïve Bayes pada Universitas XYZ” (12, 2020)[45]	(Nurhayati, Nuraeny Septianti, Nani Retnowaty, Arief Wibowo ;2020)	Pada dunia Pendidikan, menentukan kelulusan adalah hal yang sangat penting baik bagi dosen maupun mahasiswa melalui mata kuliah yang telah ditempuh tiap semester. Data ini merupakan data yang sangat penting karena dapat menjadi tolak ukur prediksi.	Penelitian ini, akan membuat prediksi yang dapat membantu program studi dalam menentukan kelulusan yang tepat waktu bagi mahasiswa	Prediksi dilakukan dengan menggunakan metode CRISP DM serta algoritma <i>naïve bayes</i> dengan hasil akurasi yang relatif baik yaitu sebesar 80,95%.

Penelitian pertama, kedua dan ketiga akan menjadi jurnal acuan pada penelitian ini digunakan untuk mempelajari algoritma *Long Short Term Memory* secara lebih lanjut. Dimana pada penelitian pertama dilakukan oleh *Yuepeng Liu* dan *Mengfei Zhang* pada tahun 2020 dengan judul “*Heart disease prediction based on random forest and LSTM*” dan dilakukan dengan

menggunakan dataset *Framingham, Heart Study*. Permasalahan yang dibahas pada penelitian ini adalah mengenai penyakit jantung yang telah menjadi ancaman dan akibat dari banyak kematian diseluruh dunia. Yang menurut *The Global Report*, resiko penyakit jantung seharusnya dapat dicegah dan dikontrol dengan mudah. Dan dengan memanfaatkan teknologi komputer serta penggunaan *machine learning* yang tepat, akan dapat membantu para tenaga kerja medis untuk mendiagnosa pasien penderita penyakit jantung dengan lebih mudah dan lebih cepat. Pada penelitian ini, dapat disimpulkan bahwa algoritma LSTM dapat menghasilkan akurasi sampai 84.5%. Namun hasil prediksi terbaik dengan tingkat akurasi sebesar 88% dicapai dengan menggunakan algoritma LSTM yang dikombinasikan dengan *random tree*[15]

Lalu penelitian kedua adalah “Prediksi Harga Beras Menggunakan Metode *Multilayer Perceptron (MLP)* dan *Long Short Term Memory (LSTM)*” yang dilakukan oleh Steven Sen, Dedy Sugiarto, Abdul Rochman pada tahun 2020. Permasalahan yang menjadi fokus pada penelitian ini adalah pentingnya ketersediaan kebutuhan pokok dimasa pandemic Covid19 salah satunya adalah beras. Sehingga pada penelitian ini dilakukan prediksi harga beras yang merupakan data *time series* dengan menggunakan pendekatan *Machine Learning* yaitu *Multilayer Perceptron (MLP)* dan *Long Short Term Memory (LSTM)*. Dan hasil dari penelitian ini adalah algoritma LSTM akan lebih akurat dalam memprediksi *trend* / harga beras jika dilihat dari rentang dengan harga aktual serta nilai RMSE-yang dihasilkan LSTM lebih kecil dibandingkan MLP.[42]

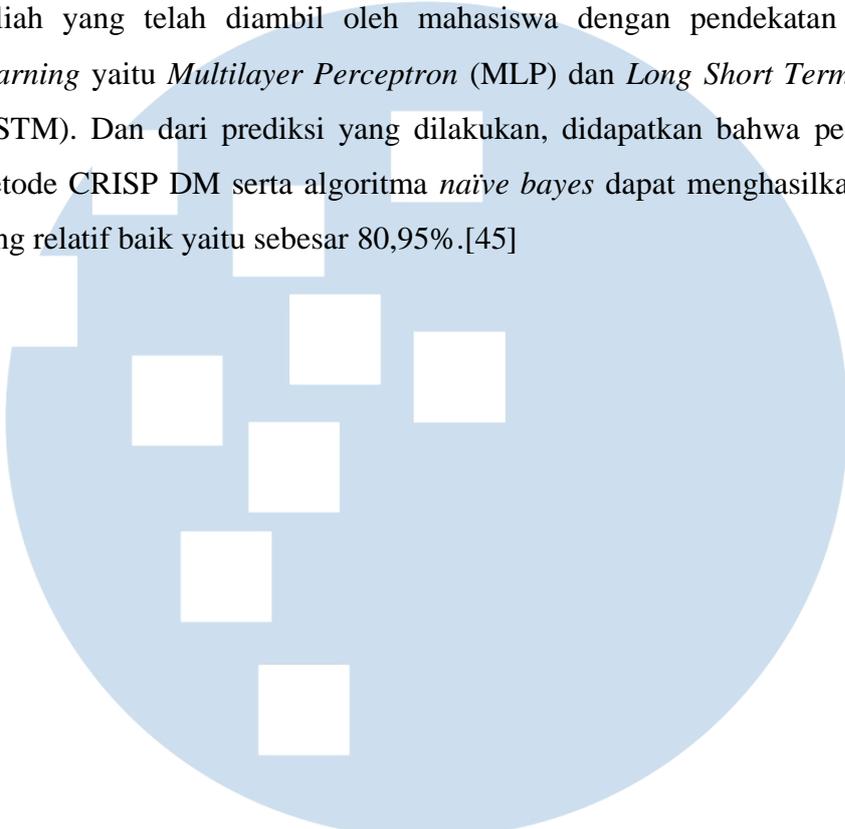
Dan untuk penelitian ketiga dengan judul “*A stacked LSTM for atrial fibrillation prediction based on multivariate ECGs*” yang dilakukan oleh Le Sun dan Yukang Wang pada tahun 2020 ini berfokus kepada permasalahan mengenai sedikitnya penelitian yang membahas mengenai deteksi AF (*Atrial Fibrillation*) dimana sebenarnya dengan melakukan deteksi AF, dapat membantu untuk mengetahui penyakit jantung sejak dini. Untuk penelitian ini, akan digunakan *Stacked RNN* yang terdiri dari LSTM bertumpuk untuk

memprediksi AF, yang disebut SLAP. Model ini dapat secara efektif menghindari ledakan gradien dan ledakan gradien RNN biasa dan mempelajari fitur-fiturnya dengan lebih baik. Setelah dilakukannya penelitian ini, ditemukan bahwa eksperimen komprehensif yang dilakukan berdasarkan dua kumpulan data public ini menunjukkan akurasi sebesar 92% dan f-score sebesar 92% dari prediksi AF, yang lebih baik daripada arsitektur deteksi AF lainnya seperti RNN dan LSTM[43].

Sementara penelitian ketiga dan keempat digunakan untuk menjadi acuan dalam mengimplementasikan metode CRISP – DM yang juga merupakan metode analisis yang digunakan pada penelitian ini. Penelitian ketiga dengan judul “*Prediction for Coronary heart Disease*” yang dilakukan oleh Neto. C pada tahun 2021. Dimana permasalahan yang menjadi fokus pada penelitian ini adalah penggunaan metode CRISP - DM dalam membangun model – model algoritma yang membantu dalam prediksi penyakit jantung. Penelitian ini dilakukan dengan membandingkan beberapa algoritma *machine learning*, diantaranya *logistic regression*, *k-nearest neighbor*, *ANN*, *SVM*, *Naïve bayes*, *Decision tree* dan juga *random forest*. Dari prediksi yang dilakukan, didapatkan kesimpulan bahwa CRISP - DM memegang peranan penting dalam bidang kesehatan, karena hasil analisa dapat digunakan untuk meningkatkan kualitas hidup pasien dan juga dapat menyelamatkan nyawa seseorang. Studi ini menunjukkan bahwa CRISP-DM memiliki peran yang sangat penting dalam memprediksi risiko penyakit jantung koroner dari faktor yang berbeda - beda[44].

Penelitian terakhir dengan judul “Prediksi Tingkat Kelulusan Tepat Waktu Mahasiswa Menggunakan Algoritma *Naïve Bayes* pada Universitas XYZ” oleh Nurhayati, Nuraeny Septianti, Nani Retnowaty, Arief Wibowo pada tahun 2020 memiliki permasalahan utama yaitu menentukan ketepatan waktu kelulusan yang merupakan aspek paling penting dalam dunia pendidikan tidak hanya bagi mahasiswa namun juga bagi para tenaga pendidik. Sehingga penelitian ini memberikan solusi dengan membuat sebuah prediksi yang dapat

menentukan ketepatan waktu kelulusan mahasiswa sesuai dengan data mata kuliah yang telah diambil oleh mahasiswa dengan pendekatan *Machine Learning* yaitu *Multilayer Perceptron* (MLP) dan *Long Short Term Memory* (LSTM). Dan dari prediksi yang dilakukan, didapatkan bahwa penggunaan metode CRISP DM serta algoritma *naïve bayes* dapat menghasilkan akurasi yang relatif baik yaitu sebesar 80,95%. [45]



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA