

2. LANDASAN TEORI

2.1 Tinjauan Pustaka

2.1.1 *Obstructive Sleep Apnea*

Obstructive sleep apnea (OSA) adalah gangguan tidur dimana seseorang mengalami perhentian pernapasan beberapa saat dan terjadi beberapa kali selama tidur. OSA sering didefinisikan dengan terjadinya mengantuk di siang hari, dengkur keras, perhentian bernapas, atau terbangun karena rasa tercekik atau sesak setidaknya lima kali kejadian pernapasan obstruktif setiap jam dari tidur (Epstein et al., 2009). OSA yang tidak diatasi dalam jangka panjang dapat menyebabkan penyakit kardiovaskular, gangguan metabolisme, gangguan kognitif, dan depresi (Osman et al., 2018). Standar emas yang digunakan untuk mendiagnosa OSA adalah *polysomnography* (PSG). PSG dilakukan dengan melakukan perekaman semalaman pada parameter seperti: aliran udara, saturasi oksigen darah, gerakan dada dan perut, mendengkur, elektrokardiografi, elektroensefalografi, elektrookulografi, elektromiografi permukaan otot *mylohyoid* dan *tibialis anterior* (Bernardini et al., 2021). PSG biasa di rumah sakit dengan memasang berbagai alat di berbagai bagian tubuh seperti kepala, jantung, jari, dan kaki.

2.1.2 Anamnesis

Anamnesis merupakan kumpulan informasi subyektif yang diperoleh dari apa yang dipaparkan oleh klien terkait dengan masalah kesehatan yang menyebabkan pasien melakukan kunjungan ke pelayanan kesehatan. Anamnesis diperoleh dari komunikasi aktif antara perawat dan pasien atau keluarga pasien. Anamnesis yang baik untuk seorang dewasa mencakupi keluhan utama, informasi mengenai kelainan yang dialami sekarang, riwayat penyakit terdahulu, riwayat keluarga, dan informasi mengenai keadaan setiap sistem tubuh pasien (Niman, 2013). Dalam anamnesis OSA terdapat beberapa kuesioner yang telah dirancang untuk melakukan skrining OSA salah satunya adalah STOP-BANG *questionnaire* dan *Epworth Sleepiness Scale*.

a. STOP-BANG *questionnaire*

Chung et al. (2016) menjabarkan bahwa STOP-BANG *questionnaire* merupakan akronim dari pertanyaan ya atau tidak yang terdiri dari *Snoring, Tiredness, Observed apnea, high blood Pressure, BMI, Age, Neck circumference*, dan *male Gender*. Setiap pertanyaan yang dijawab ya akan mendapatkan satu poin sehingga total poin akan berada pada rentan nol sampai delapan.

b. *Epworth Sleepiness Scale*

Johns (1991) mendeskripsikan *Epworth Sleepiness Scale* (ESS) sebagai tolak ukur untuk mengetahui apakah seseorang mengalami kelelahan berlebih di siang hari. ESS menggunakan skala 0 sampai 3 untuk mengukur kemungkinan seseorang tertidur ketika beraktivitas di siang hari (0 tidak mungkin tertidur, 1 kecil kemungkinan tertidur, 2 mungkin tertidur, 3 sangat mungkin tertidur). Tabel 6.1 menunjukkan daftar kegiatan yang ditanyakan dalam ESS. Hasil skala tersebut lalu ditotal menjadi sebuah skor yang jika skor dibawah 10 menandakan cukup tidur, diantara 10 dan 16 menandakan kelelahan berlebihan di siang hari, dan diatas 16 menandakan mengantuk yang membahayakan.

2.1.3 Antropometri

Casadei & Kiel (2021) mendefinisikan antropometri sebagai pengukuran kuantitatif non-invasif pada tubuh. Unsur inti antropometri adalah tinggi badan, berat badan, lingkar kepala, indeks massa tubuh (BMI), lingkar tubuh untuk menilai adipositas (pinggang, pinggul, dan anggota badan), dan ketebalan lipatan kulit.

2.1.4 *Machine Learning*

Machine Learning adalah bagian dari *Artificial Intelligence* yang digunakan untuk mengembangkan program komputer. *Machine Learning* dipandu oleh kumpulan data yang luas menggunakan algoritma dan statistik tertentu (Singh et al., 2023). *Machine Learning* membantu dalam melacak data dengan memprediksi pola-pola yang muncul. Beberapa teknik dalam *Machine Learning* menggunakan kumpulan data besar, membentuk pola berdasarkan data masa lalu, dan memperkirakan masa depan. Pembelajaran mesin memungkinkan komputer untuk melakukan pemodelan berdasarkan pengalaman dan memprediksi hasil di masa depan. *Machine Learning* memiliki tiga hal utama: kelas tugas, ukuran kinerja, dan pengalaman yang terdefinisi dengan baik.

2.1.4.1 *Support Vector Machine (SVM)*

Support Vector Machine (SVM) adalah algoritma klasifikasi yang efektif dalam menangani tugas klasifikasi biner maupun multi kelas. SVM bekerja dengan mencari batas keputusan optimal yang memaksimalkan margin antara kelas-kelas yang berbeda. SVM juga dapat mengatasi masalah data yang tidak linier melalui penggunaan kernel untuk mentransformasi ruang fitur. Dalam pengenalan pola, fungsi diskriminan linier dalam ruang berdimensi n dapat dinyatakan sebagai: $g(x) = \omega \cdot x + b$. Persamaan *hyperplane* klasifikasi kemudian dapat dituliskan sebagai $g(x) = \omega \cdot x + b = 0$. Pada kasus yang dapat dipisahkan secara linier, fungsi diskriminan $g(x)$ dinormalisasi sehingga semua

sampel pelatihan memenuhi $|g(x)| \geq 1$, bahkan untuk sampel yang berada jauh dari permukaan klasifikasi $|g(x)| \geq 1$. Dengan demikian, interval kelas setara dengan $2\|\omega\|$, sehingga interval ini menjadi setara dengan $\|\omega\|$ atau $\|\omega\|^2$. Agar permukaan klasifikasi dapat mengklasifikasikan semua sampel dengan benar, perlu dipenuhi:

$$y_i[(\omega \cdot x_i) + b] - 1 \geq 0, \quad i = 1, 2, \dots, n \quad (2.1)$$

memenuhi persamaan (2.1 dan membuat $\|\omega\|^2$ seminimal mungkin akan menghasilkan permukaan klasifikasi yang optimal. Titik-titik pada *hyperplane* disebut vektor pendukung (*support vectors*), yang mendukung permukaan klasifikasi optimal (Jandik & Makhija, 2005).

2.1.4.2 **Random Forest**

Random Forest (RF) adalah kombinasi dari banyak *decision tree*. Setiap *decision tree* dibangun berdasarkan set sampel pelatihan dan variabel acak. Variabel acak yang sesuai dengan *tree* ke- k dilambangkan sebagai θ_k , dan variabel-variabel ini bersifat independen serta terdistribusi identik, yang menghasilkan pengklasifikasi $h(x, \theta_k)$, dimana x adalah vektor input. Setelah menjalankan proses ini sebanyak k kali, didapatkan urutan pengklasifikasi $\{h_1(x), h_2(x), \dots, h_k(x)\}$ dan menggunakan urutan ini untuk membentuk sistem model klasifikasi. Hasil akhir dari sistem ini ditentukan dengan voting mayoritas biasa. Fungsi keputusan ini adalah:

$$H(x) = \arg \max_Y \sum_{i=1}^k I(h_i(x) = Y) \quad (2.2)$$

dimana $H(x)$ adalah kombinasi dari model klasifikasi, h_i adalah model *decision tree* tunggal, Y adalah variabel output, dan $I(\cdot)$ adalah fungsi indikator. Untuk variabel input yang diberikan, setiap *decision tree* dalam hutan memiliki hak untuk memberikan suara dan memilih hasil klasifikasi terbaik (Liu et al., 2012).

2.1.4.3 **K-Nearest Neighbors (KNN)**

K-Nearest Neighbors (KNN) adalah algoritma yang berbasis pada konsep bahwa data dengan fitur serupa cenderung berada dalam kelas yang sama. KNN mengklasifikasikan suatu data berdasarkan mayoritas kelas dari k tetangga terdekatnya dalam ruang fitur. Misalkan terdapat data x_i dan x_j , kedekatan dua data tersebut didefinisikan sebagai $d(x_i, x_j)$ pada persamaan (2.3 yang merupakan jarak Euclidian dimana N merupakan jumlah fitur yang dimiliki kedua data (Mitchel, 1997).

$$d(x_i, x_j) = \sqrt{\sum_{n=1}^N (x_{i,n} - x_{j,n})^2} \quad (2.3)$$

2.1.4.4 Logistic Regression

Logistic Regression adalah sebuah algoritma klasifikasi di mana variabel responsnya bersifat kategorikal. Algoritma ini digunakan untuk menemukan hubungan antara fitur-fitur dan probabilitas dari hasil tertentu. Dalam klasifikasi biner, XXX merupakan fitur-fitur dan YYY adalah output yang bisa bernilai 0 atau 1. Logistic regression menggunakan fungsi yang disebut fungsi *logit*, yang membantu menggambarkan hubungan antara variabel dependen dan variabel independen dengan memprediksi probabilitas terjadinya suatu peristiwa. *Odds* adalah rasio antara probabilitas keberhasilan (P) dan probabilitas kegagalan (1-P), sedangkan *logit* adalah log dari rasio *odds* tersebut. Dengan demikian, persamaan *logit* dapat dinyatakan sebagai persamaan (2.4). Selanjutnya, untuk mendapatkan persamaan probabilitas (persamaan (2.5)), dilakukan antilog pada persamaan (2.4) (Abhigyan, 2020).

$$\log(Y) = \ln\left(\frac{P}{1-P}\right) = a + b \cdot X \quad (2.4)$$

$$P = \frac{e^{a+b \cdot X}}{1 + e^{a+b \cdot X}} \quad (2.5)$$

2.1.5 Matrix Evaluasi

2.1.5.1 Confusion Matrix

Confusion Matrix merupakan salah satu cara paling populer untuk melakukan evaluasi pada model *machine learning* klasifikasi (Singh et al., 2021). *Confusion Matrix* menghitung jumlah prediksi dan nilai sesungguhnya dari klasifikasi. Umumnya *confusion matrix* digunakan pada klasifikasi biner (dua *class*) saja tetapi *confusion matrix* juga dapat digunakan pada klasifikasi dengan *multiclass*. *Confusion Matrix* terdiri dari empat elemen utama seperti terlihat pada Tabel 2.1:

Tabel 2.1
Confusion Matrix

		Prediksi	
		Positif	Negatif
Sebenarnya	Positif	TP	FN
	Negatif	FP	TN

- *True Positive* (TP): Jumlah prediksi benar untuk kelas positif.
- *True Negative* (TN): Jumlah prediksi benar untuk kelas negatif.
- *False Positive* (FP): Jumlah prediksi salah untuk kelas positif.
- *False Negative* (FN): Jumlah prediksi salah untuk kelas negatif.

2.1.5.2 Akurasi

Akurasi adalah ukuran seberapa sering model membuat prediksi yang benar yang dihitung sebagai rasio jumlah prediksi yang benar dengan total jumlah prediksi seperti terlihat pada persamaan (2.6 (Singh et al., 2021).

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN} \quad (2.6)$$

2.1.5.3 Spesifisitas

Spesifisitas mengukur seberapa baik model dapat mengidentifikasi semua data negatif. Spesifisitas adalah rasio jumlah prediksi benar untuk kelas negatif terhadap jumlah total data negatif yang sebenarnya seperti terlihat pada persamaan (2.7 (Singh et al., 2021)..

$$Spesifisitas = \frac{TN}{TN + FP} \quad (2.7)$$

2.1.5.4 Sensitivitas

Sensitivitas, juga dikenal sebagai *recall* atau *true positive rate*, adalah ukuran seberapa baik model dapat mengidentifikasi semua data positif. Sensitivitas adalah rasio jumlah prediksi benar untuk

kelas positif terhadap jumlah total data positif yang sebenarnya seperti terlihat pada persamaan (2.8) (Singh et al., 2021).

$$\text{Sensitivitas} = \frac{TP}{TP + FN} \quad (2.8)$$

2.1.5.5 Nilai Prediktif Positif

Nilai prediktif positif, atau *precision*, mengukur seberapa andal prediksi positif dari model. Ini adalah rasio jumlah prediksi benar untuk kelas positif terhadap jumlah total prediksi positif yang dibuat oleh model seperti terlihat pada persamaan (2.9) (Singh et al., 2021).

$$\text{Nilai Prediktif Positif} = \frac{TP}{TP + FP} \quad (2.9)$$

2.1.5.6 Nilai Prediktif Negatif

Nilai prediktif negatif mengukur seberapa andal prediksi negatif dari model. Nilai prediktif negatif adalah rasio jumlah prediksi benar untuk kelas negatif terhadap jumlah total prediksi negatif yang dibuat oleh model seperti terlihat pada persamaan (2.10) (Singh et al., 2021).

$$\text{Nilai Prediktif Negatif} = \frac{TN}{TN + FN} \quad (2.10)$$

2.1.5.7 F1 Score

F1 Score memberikan gambaran keseimbangan antara *precision* (nilai prediktif positif) dan *recall* (sensitivitas) seperti terlihat pada persamaan (2.11) (Singh et al., 2021).

$$F1 \text{ Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (2.11)$$

2.2 Tinjauan Studi

2.2.1 *Screening for Obstructive Sleep Apnea Risk by Using Machine Learning Approaches and Anthropometric Features*

Penelitian yang dilakukan Tsai et al. (2022) mengangkat masalah diagnosis OSA menggunakan PSG di laboratorium selama tidur sulit dilakukan karena memakan waktu dan tenaga. Penelitian ini melakukan skrining OSA dengan menggunakan data antropometri pasien dan data tidur pasien menggunakan PSG. Metode yang digunakan dalam penelitian ini terdapat enam macam model *machine learning* seperti *logistic regression* (LR), *k-nearest neighbors* (kNN), *naïve Bayes* (NB), *random forest* (RF), *support vector machine* (SVM), dan *extreme gradient boosting* (XGBoost). Model RF menjadi yang memiliki performa akurasi paling tinggi yaitu 84.74% untuk OSA sedang hingga parah dan 72.61% untuk OSA parah. Fitur yang dianggap paling berpengaruh untuk menentukan risiko OSA adalah lemak perut.

2.2.2 *Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning*

Ramesh et al. (2021) dalam penelitian ini mengangkat masalah bahwa standar tradisional untuk diagnosa OSA yaitu PSG memiliki akurasi yang tinggi tetapi tidak praktis untuk pasien karena perlu menginap semalaman dan memiliki biaya yang tinggi. Penelitian ini menggunakan input data yang cukup lengkap yaitu anamnesis, antropometri, tes darah, dan PSG. Hasil penelitian ini memperoleh hasil yang cukup memuaskan dengan model SVM mendapatkan akurasi tertinggi (68.06%), *sensitivity* (88.76%), dan *F1-score* (75.96%). Model *logistic regression* mendapatkan *specificity* tertinggi (40.74%), sementara model *random forest* mendapatkan nilai tertinggi pada nilai prediksi positif (66.36%) dan nilai prediksi negatif (73.33%).

2.2.3 *Clinical Trial Enrollment Enrichment in Resource-Constrained Research Environments: Multivariable Apnea Prediction (MAP) Index in SCIP-PA Trial*

Penelitian yang dilakukan oleh Yang et al. (2018) mengangkat masalah uji klinis OSA yang kurang efisien dalam sumber daya karena memerlukan peralatan khusus, mahal, dan memakan waktu. Penelitian ini menggunakan metode *Multivariable Apnea Prediction* (MAP) untuk melakukan skrining. MAP menggunakan *logistic regression* sebagai model untuk memprediksi pasien yang memiliki kemungkinan OSA. Hasil dari penelitian ini menunjukkan *sensitivity* yang cukup baik yaitu 83.6 % dan *specificity* 46.4%. Namun, secara keseluruhan *sensitivity* dan *specificity* untuk pria lebih bagus daripada wanita membuat adanya gender bias untuk skrining OSA.