

## **BAB 2**

### **LANDASAN TEORI**

#### **2.0.1 Polycystic Ovarian Syndrome (PCOS)**

PCOS adalah kontributor signifikan terhadap infertilitas ovulasi dan berdampak pada lima hingga 20% wanita usia subur [1],[21]. Penyakit PCOS ini kerap dikatakan sebagai penyakit kelainan hormon yang banyak ditemukan pada wanita subur (15-44 tahun) [22]. Hampir sekitar 70% dari semua yang menderita penyakit PCOS ini tidak terdiagnosis [23]. Masih minim sekali tingkat kepedulian wanita akan penyakit ini, padahal PCOS dapat memengaruhi kesuburan serta kemampuan dalam memiliki keturunan dalam jangka panjang [22].

Secara Global, penyakit PCOS ini sudah diperkirakan sebesar 3,4% yang terkena penyakit ini. Di Amerika Serikat juga dipengaruhi sekitar lima juta wanita yang berusia produktif terkena PCOS, sekitar 6,5% hingga 8% prevalensi PCOS di laporkan di Eropa, 15,5% pasien PCOS di Thailand, dan 8,7% prevalensi dilaporkan di Australia [5]. Lalu untuk Indonesia masih belum memiliki data resmi terkait dengan prevalensi nasional POCS, tetapi dalam studi di RS Cipto Mangunkusumo telah diidentifikasi terdapat sekitar 105 pasien penderita POCS. Dari jumlah pasien tersebut memiliki beberapa keluhan, yaitu keluhan oligo atau amenore dengan 94,2% pasien dan keluhan mengalami hirsutisme dengan 32,4%. Terdapat sebanyak 45,7% pasien mayoritas memiliki rentang usia sekitar 26-30 tahun [5].

#### **A Penyebab Polycystic Ovarian Syndrome (PCOS)**

Penyebab dari penyakit PCOS masih belum sepenuhnya diketahui, akan tetapi terdapat beberapa faktor yang diduga dapat terkait dengan PCOS, yaitu:

1. Kelebihan hormon insulin menjadi salah satu faktor terkait dengan PCOS yaitu hormon yang akan menurunkan kadar gula yang terdapat pada darah. Dengan meningkatnya kadang insulin membuat peningkatan produksi hormon androgen serta akan menurunkan sensitivitas pada tubuh terhadap insulin.
2. Faktor genetik bisa menjadi salah satu faktor dikarenakan beberapa pasien penderita PCOS memiliki anggota keluarga yang juga penderita PCOS.

## **B Gejala Polycystic Ovarian Syndrome (PCOS)**

Gejala dari PCOS bisa saja timbul saat wanita mengalami haid untuk pertama kalinya di masa pubertas, akan tetapi masih ada beberapa gejala yang menjadi gambaran bahwa terjangkit penyakit PCOS [2].

1. Gangguan menstruasi terjadi dengan periode menstruasi yang tidak teratur atau berkepanjangan.
2. Memiliki kadar hormon androgen yang meningkat dengan PCOS pada wanita akan menimbulkan gejala fisik seperti pria. Gejala ini bisa seperti menumbuhnya rambut yang lebat di wajah dan juga bagian tubuh (hirsutisme), dan lain-lain
3. Penderita dapat ditemukan terdapat kantong-kantong kista yang terdapat di sekitar ovarium atau sel telur, biasa disebut sebagai kista ovarium.
4. Penderita akan mengalami penggelapan pada kulit terutama di area lipatan (leher, selangkangan, dan lain-lain.).

## **C Pencegah dan Pengobatan Polycystic Ovarian Syndrome (PCOS)**

Untuk pencegahan dari penyakit PCOS sendiri masih belum seutuhnya diketahui penyebab utamanya, sehingga untuk melakukan pencegahannya bisa dengan melakukan pengendalian dari faktor risiko agar PCOS dapat dicegah. Untuk mengobati PCOS terdapat beberapa cara antara lain:

1. Melakukan perubahan gaya hidup
2. Olahraga secara teratur
3. Memperbanyak konsumsi serat
4. Mengonsumsi pil kontrasepsi (pil KB)
5. Mengonsumsi obat anti androgen agar hormon androgen yang berlebihan sebelumnya dapat disesuaikan dengan jumlah yang seharusnya.

## 2.0.2 K-Nearest Neighbor Classifier

Menurut salah satu situs web Trivusi, Algoritma K-Nearest Neighbors (KNN) merupakan salah satu algoritma *machine learning* yang sederhana dan juga dapat diterapkan dengan mudah. Algoritma ini dapat membantu untuk menyelesaikan masalah dalam regresi dan juga klasifikasi. KNN menjadi salah satu algoritma jenis *supervised learning* [24]. Dalam melakukan perhitungan dalam KNN juga dapat dimulai dengan menentukan jumlah dari tetangga yang nantinya akan diperhitungkan atau sebagai acuan saat menentukan kelas dari data baru. KNN juga menjadi salah satu algoritma *lazy learning* dikarenakan tidak membutuhkan banyak data untuk bisa dilakukan proses pembelajarannya. Knn juga akan dilakukan dengan mencari kelompok pada K objek dalam *data training* yang terdekat (mirip) dengan objek pada data baru atau *data testing* [25].

Algoritma KNN bekerja berdasarkan pada jarak terpendek dari *sample* uji ke *sample* latih untuk menentukan KNNnya. Algoritma KNN memiliki keunggulan yaitu mudah dimengerti, dipelajari, serta dalam melakukan pelatihannya cukup sederhana, cepat, serta dianggap efektif saat melakukan pelatihan menggunakan data yang cukup besar. Namun, KNN juga memiliki kelemahan yaitu pada nilai  $k$ fold. Terdapat langkah-langkah dalam melakukan pembelajaran menggunakan algoritma KNN, antara lain [20]:

1. Memastikan parameter K
2. Menentukan jarak antara *data testing* dan *data training* Dalam melakukan perhitungan jarak, KNN memiliki dua cara yaitu dengan jarak Euclidean dan jarak Manhattan.
3. Mengurutkan jarak yang akan terbentuk
4. Memastikan jarak terdekat sampai pada urutan K
5. Memasukkan kelas yang sesuai
6. Menentukan jumlah kelas dari tetangga terdekat dan menentukan kelas sebagai data yang nantinya akan dilakukan evaluasi.

### A Menghitung Jarak Kemiripan (K)

Dalam melakukan perhitungan jarak kemiripan dalam nilai K perlu dilakukan dengan 2 cara yaitu dengan Manhattan/City Block distance ( $p=1$ ), seperti

pada persamaan (1).

$$d(x,y) = \sum_{i=1}^n |x_i - y_i| \quad (2.1)$$

Lalu untuk perhitungan selanjutnya dengan Euclidean distance ( $p=2$ ), seperti persamaan (2).

$$d(x,y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (2.2)$$

dimana:

$x_i$  = sample data / *data training*

$y_i$  = data uji / *data testing*

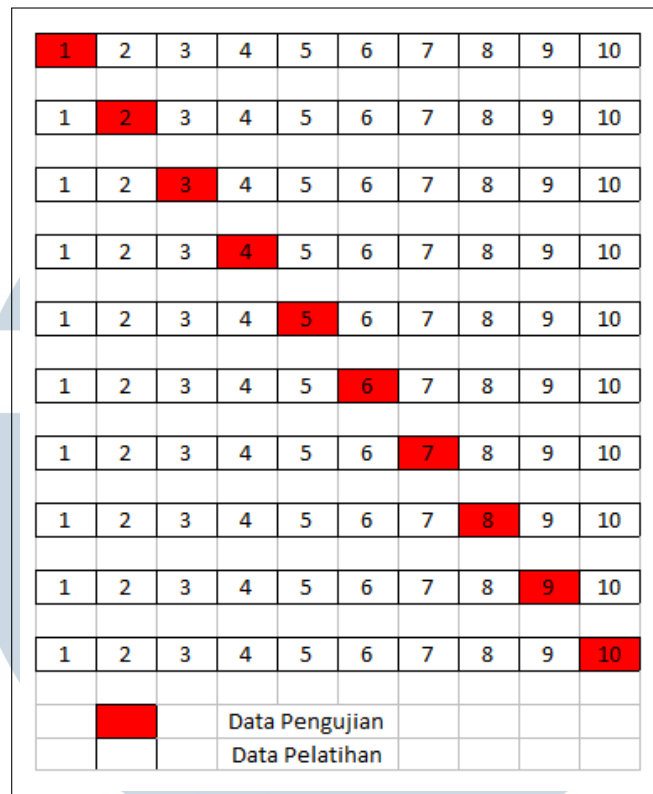
$i$  = variabel data

$n$  = dimensi data

### 2.0.3 K-Fold Cross Validation

*Cross Validation* merupakan teknik untuk menilai hasil dari statistik analisis akan menggeneralisasikan kumpulan data independen [26]. Teknik ini utamanya digunakan untuk melakukan prediksi suatu model dan memperkirakannya seberapa akurat sebuah model prediktif ketika dijalankan dalam praktik yang sebenarnya [26]. Dalam menentukan nilai K pada KNN dengan optimasi parameter, *K-Fold Cross Validation* yang akan menjadi salah satu penentu untuk mengetahui nilai K yang paling optimal [27]. *K-Fold* merupakan sistem yang dijalankan dengan melakukan perulangan secara acak pada atribut yang dimasukkan sehingga sistem yang akan menguji untuk beberapa atribut input yang acak tersebut [27].

Saat melakukan *cross validation*, akan dikerjakan sebanyak K kali untuk proses *training* dan *testing*. Dalam menjalankan *cross validation* atau validasi silang sebanyak 10 kali (*10-Fold*) terbukti menghasilkan performa algoritma yang lebih stabil sehingga sudah banyak peneliti yang menggunakannya [28]. Representasi 10 folds cross validation dapat dilihat pada Gambar 2.1[29].



Gambar 2.1. Skema 10-Fold cross validation

## 2.0.4 Confussion Matrix

*Confussion Matrix* merupakan salah satu evaluasi model yang akan memberikan informasi dari perbandingan hasil klasifikasi yang dilakukan oleh sistem (model) dengan hasil klasifikasi yang sebenarnya [30]. *Confussion matrix* merupakan sebuah tabel yang memiliki 4 kombinasi hasil sebagai pengukuran kerja yaitu True Positive, True Negative, False Positive, dan False Negative [31].

Tabel 2.1. Confusion matrix

		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

Keempat istilah dari isi Tabel *Confussion Matrix* memiliki makna masing-masing, yaitu

1. *True Negative* merupakan hasil data negatif yang diprediksi benar.
2. *True Positive* merupakan hasil data positif yang diprediksi benar.
3. *False Positive* merupakan hasil data negatif akan tetapi diprediksi sebagai data positif.
4. *False Negative* merupakan hasil data positif akan tetapi di prediksi sebagai data negatif.

Untuk memudahkan memahami dan mengingat penjelasan diatas adalah jika diawali dengan kata *True* diprediksi benar baik terjadi maupun tidak terjadi, jika diawali dengan *False* diprediksi salah, dan untuk positif dan negatif merupakan hasil prediksi yang didapatkan dari model [30]. Jika sudah mendapatkan nilai dari *Confussion Matrix*, maka bisa dilakuakn perhitungan selanjutnya yaitu terdapat *Recall*, *Accuracy*, *Precision*, dan *F-1 Score*.

Dalam melakukan perhitungan *recall* akan memperlihatkan hasil persentase keberhasilan model untuk menemukan kembali informasi [32].

$$Recall = \frac{TP}{(TP + FN)} \quad (2.3)$$

Dalam melakukan perhitungan *accuracy* akan memberikan persentase hasil dari seberapa akurat menggunakan model prediksi dengan benar [32].

$$Accuracy = \frac{(TP + TN)}{(TP + FN + TN + FP)} \quad (2.4)$$

Dalam melakukan perhitungan *Precision* akan menggambarkan besar hasil akurasi dari model yang diberikan antara data yang diminta dengan hasil prediksi [32].

$$Precision = \frac{TP}{(TP + FP)} \quad (2.5)$$

Dalam melakukan perhitungan F-1 Score akan memberikan hasil dari rata-rata perbandingan hasil *pression* dan *recall* [32].

$$F-1Score = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (2.6)$$