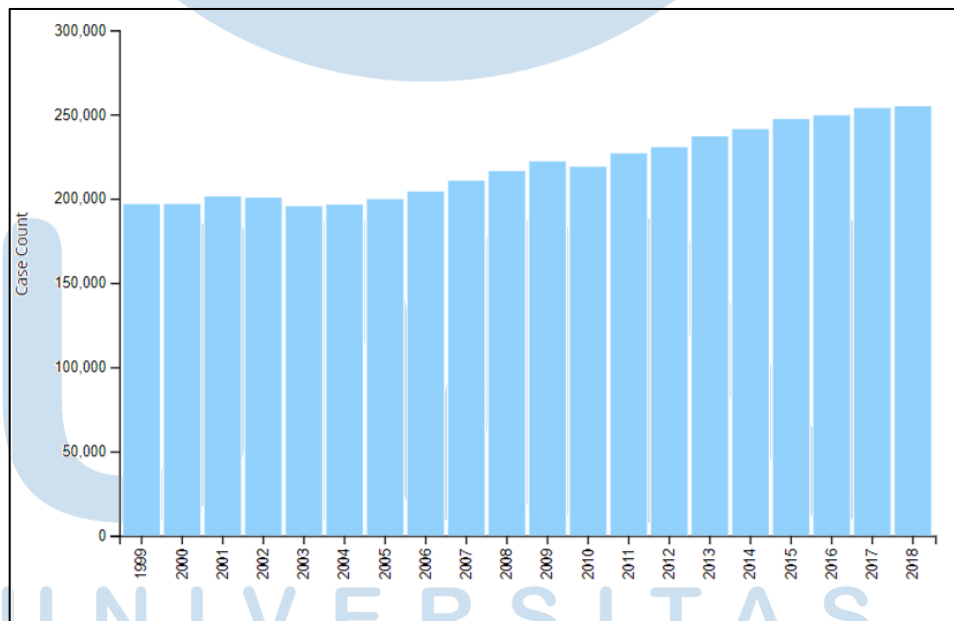


# BAB I

## PENDAHULUAN

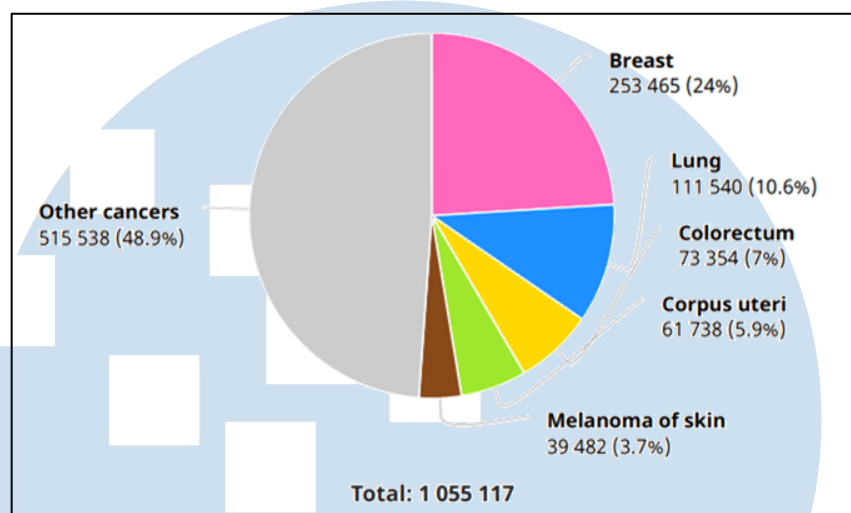
### 1.1 Latar Belakang

Kanker payudara merupakan salah satu penyakit kanker yang memiliki kasus terbanyak dan terjadi hampir di seluruh dunia. Setiap tahun, pasti ada kasus kematian yang disebabkan oleh kanker payudara. Menurut salah satu artikel di *newsroom World Health Organization (WHO)*, di tahun 2020 diketahui bahwa ada sekitar 2,3 juta wanita di dunia yang terdiagnosa mengidap kanker payudara dengan total kasus kematian mencapai 685.000 jiwa. Hingga pada akhir tahun 2020, dalam 5 tahun terakhir ada total sekitar 7.8 juta wanita yang terdiagnosa menderita kanker payudara [1]. Kanker payudara juga menjadi penyakit serius di negara *United State of America (USA)*. Kasus kanker payudara yang ada di negara tersebut menunjukkan adanya peningkatan trend dari tahun ke tahun seperti yang ditunjukkan pada gambar 1.1 [2].



**Gambar 1.1 Trend Kanker Payudara Hingga 1999-2018 [2]**

Dari sumber data Globocan 2020, diketahui bahwa pada tahun 2020 di USA terdapat 253.465 kasus baru kanker payudara dan semua yang terpapar adalah wanita seperti yang terlampir pada gambar 1.2 [3].



**Gambar 1.2 Chart Jenis Kanker yang Menyerang Wanita 2020 [3]**

Di tahun 2020, kanker payudara juga merupakan penyakit kanker dengan jumlah kasus paling banyak di USA dan memiliki kasus kematian ketiga tertinggi setelah penyakit kanker paru-paru dan kanker pankreas dengan jumlah 42.617 kematian di tahun 2020 seperti yang terlampir pada gambar 1.3 [3].

Cancer	New cases				Deaths			
	Number	Rank	(%)	Cum.risk	Number	Rank	(%)	Cum.risk
Breast	253 465	1	11.1	9.81	42 617	3	7.0	1.36
Lung	227 875	2	10.0	4.05	138 225	1	22.6	2.24
Prostate	209 512	3	9.2	9.29	32 438	5	5.3	0.71
Colon	101 809	4	4.5	1.79	37 930	4	6.2	0.56
Melanoma of skin	96 445	5	4.2	1.82	7 201	18	1.2	0.11
Bladder	80 617	6	3.5	1.27	18 130	10	3.0	0.19
Non-Hodgkin lymphoma	73 652	7	3.2	1.34	20 858	8	3.4	0.26
Kidney	69 569	8	3.0	1.43	14 589	13	2.4	0.24
Corpus uteri	61 738	9	2.7	2.60	11 460	16	1.9	0.37
Leukaemia	61 152	10	2.7	1.09	23 753	7	3.9	0.31
Pancreas	56 654	11	2.5	0.97	47 683	2	7.8	0.77
Thyroid	52 912	12	2.3	1.19	2 161	25	0.35	0.03
Rectum	44 780	13	2.0	0.95	15 139	12	2.5	0.26
Liver	42 284	14	1.9	0.89	31 078	6	5.1	0.59
Multiple myeloma	32 119	15	1.4	0.59	13 426	15	2.2	0.19
Stomach	26 259	16	1.2	0.48	11 413	17	1.9	0.19
Brain, central nervous system	24 538	17	1.1	0.51	18 133	9	3.0	0.36
Lip, oral cavity	24 470	18	1.1	0.51	4 285	20	0.70	0.07
Ovary	23 820	19	1.0	0.91	14 359	14	2.3	0.47
Oesophagus	18 309	20	0.80	0.35	16 209	11	2.6	0.30
Cervix uteri	13 545	21	0.59	0.60	5 706	19	0.93	0.22
Oropharynx	12 775	22	0.56	0.30	3 200	22	0.52	0.06
Larynx	12 554	23	0.55	0.27	3 833	21	0.63	0.07
Testis	9 407	24	0.41	0.43	450	32	0.07	0.02
Anus	8 419	25	0.37	0.18	1 374	27	0.22	0.03
Hodgkin lymphoma	8 107	26	0.36	0.18	943	29	0.15	0.01
Vulva	6 112	27	0.27	0.21	1 487	26	0.24	0.03
Salivary glands	4 922	28	0.22	0.09	956	28	0.16	0.01
Gallbladder	4 670	29	0.20	0.08	2 301	24	0.38	0.04
Mesothelioma	3 409	30	0.15	0.05	2 538	23	0.41	0.03
Hypopharynx	2 245	31	0.10	0.05	549	31	0.09	0.01
Nasopharynx	1 898	32	0.08	0.04	915	30	0.15	0.02
Penis	1 515	33	0.07	0.05	414	34	0.07	0.01
Vagina	1 446	34	0.06	0.05	414	33	0.07	0.01
Kaposi sarcoma	1 078	35	0.05	0.02	92	35	0.02	0.00
<b>All cancer sites</b>	<b>2 281 658</b>	<b>-</b>	<b>-</b>	<b>34.06</b>	<b>612 390</b>	<b>-</b>	<b>-</b>	<b>9.18</b>

**Gambar 1.3 Data Kasus Baru dan Kematian Penyakit Kanker tahun 2020 [3]**

Menurut WHO, kanker payudara atau *breast cancer* adalah penyakit yang biasanya muncul di sel-sel lapisan *epitel*, di saluran *duktus*, dan di kelenjar yang menghasilkan susu atau disebut *lobulus*. Sebenarnya, pertumbuhan kanker pada daerah *duktus* dan *lobulus* tidak menimbulkan gejala dan resiko penyebarannya minim atau dapat dikatakan bersifat metastasis. Namun kanker yang tumbuh di *duktus* dan *lobulus* ini dapat berkembang dan menyerang ke jaringan payudara di sekitarnya sehingga inilah yang disebut kanker payudara [1].

Kanker payudara tidaklah menular dan tidak hanya menyerang kaum wanita. Menurut salah satu jurnal yang diterbitkan oleh *American Cancer Society*, dikatakan bahwa kanker payudara juga bisa menyerang pria. Namun memang penyakit ini lebih umum menyerang wanita. Biasanya penyakit kanker payudara menyerang orang-orang yang memasuki usia 55 tahun atau usia lanjut. Sekitar 5% hingga 10% penderita kanker payudara juga disebabkan oleh faktor gen atau keturunan [4].

Meskipun terpapar, ada beberapa cara yang dapat dilakukan penderita untuk dapat mengobati penyakit kanker payudara. Menurut jurnal “*Advances in Breast Cancer Radiation Therapy*”, penyakit kanker payudara dapat disembuhkan dengan beberapa pengobatan. Pengobatan yang dapat dilakukan itu seperti *surgical resection*. *Surgical resection* adalah pengobatan dengan operasi pengangkatan sebagian atau seluruh tumor. *Surgical resection* ini perlu diikuti dengan program kemoterapi, terapi radiasi, serta terapi endokrin yang dapat disesuaikan dengan tingkat keparahan penyakit payudara [5].

Seseorang yang terdeteksi menderita kanker payudara memang berpotensi untuk sembuh dengan melakukan perawatan, namun tidak menutup kemungkinan bahwa penyakit tersebut akan bertambah parah hingga beresiko pada kematian. Oleh sebab itu untuk meminimalisir terpaparnya kanker payudara dapat dilakukan upaya deteksi dini dengan melakukan prediksi menggunakan teknologi. Dalam melakukan prediksi penyakit seperti halnya kanker payudara, sudah banyak penelitian-penelitian terdahulu yang menggunakan teknologi *data mining* untuk klasifikasi dan memprediksi data. Dengan pemanfaatan *data mining* akan membantu mendiagnosis kanker lebih awal sehingga dapat segera melakukan perawatan sebelum kanker tersebut mengganas [6].

Salah satu penelitian terdahulu berjudul “*Machine learning Algorithms For Breast Cancer Prediction And Diagnosis*” [7] juga melakukan deteksi dini kanker payudara dengan menggunakan pemodelan algoritma. Penelitian menggunakan lima macam algoritma yang diantaranya adalah *Support Vector Machine* (SVM), *Random Forest* (RF), *Logistic Regression* (LR), *k-Nearest Neighbor* (KNN), dan *Decision Tree* (DT). Hasil penelitian ini, diperoleh akurasi tertinggi dari algoritma SVM sebesar 97,2%. Kemudian untuk *Random Forest* diperoleh akurasi 96,5%, *Logistic Regression* sebesar 95,8%, *k-Nearest Neighbor* sebesar 93,7%, dan *Decision Tree* sebesar 95,1%.

Penelitian berikutnya berjudul “*Diagnosis of Breast Cancer Based on Support Vector Machine and Random Forest Method*” [8] juga melakukan

prediksi kanker payudara. Penelitian ini menggunakan dua macam algoritma yaitu *Support Vector Machine (SVM)* dan *Random Forest*. Pada penelitian ini diperoleh akurasi terbaik dari algoritma SVM. Nilai akurasi dari algoritma SVM adalah 97% sedangkan *Random Forest* sebesar 96%.

Penelitian “*Predicting the Possibility of Cancer with Supervised Learning Algorithms*” [9] yang dilakukan pada tahun 2020 juga melakukan prediksi kanker payudara. Pada penelitian ini digunakan algoritma *Nearest Neighbor (NN)*, *Support Vector Machine (SVM)*, *Naive Bayes (NB)*, dan *Decision Tree (DT)*. Dari penelitian ini diperoleh akurasi tertinggi dari algoritma SVM yaitu sebesar 97,2%. Sedangkan akurasi algoritma *Nearest Neighbor* sebesar 95,1%, *Naive Bayes* sebesar 91,6%, dan *Decision Tree* sebesar 95,8%.

Dari contoh penelitian [7-9] dapat dilihat bahwa penelitian terdahulu sudah menerapkan *data mining* dengan menggunakan berbagai macam algoritma. Performa akurasi dari algoritma sendiri juga sudah baik dan angkanya sudah diatas nilai 90% terlebih pada algoritma SVM yang memiliki akurasi tertinggi dibandingkan dengan algoritma lainnya yaitu mencapai 97,2%. Namun akan lebih baik lagi apabila kita dapat mencoba memodifikasi algoritma SVM ini dengan teknik lain seperti dengan menjabarkan dari tiap *kernel*nya untuk membandingkan akurasinya.

Selain dari algoritma SVM, proses penerapan *data mining* pada algoritma tersebut juga dapat dilakukan dengan menggunakan *kernel*. Pada penelitian “*Computational Predictions for Protein Sequences of COVID-19 Virus via Machine learning Algorithms* [10]” melakukan proses *data mining* salah satunya menggunakan algoritma SVM dengan membandingkan *Kernel Linear*, *Kernel RBF*, dan *Kernel Sigmoid* dengan diperoleh akurasi *Kernel Linear* sebesar 100%, *Kernel RBF* sebesar 99.6%, dan *Kernel Sigmoid* sebesar 95.2%. Pada penelitian “*Coronavirus disease (COVID-19) detection in Chest X-Ray images using majority voting based classifier ensemble* [11]” juga melakukan perbandingan *Kernel* pada SVM dengan diperoleh hasil akurasi pada *Kernel Linear* 96.124% dan *Kernel RBF* 95.349%.

Selain itu algoritma SVM dan penggunaan *Kernel*, penelitian terdahulu ada juga yang memodifikasi algoritmanya dengan menggunakan *hyperparameter* yaitu pada penelitian “Covid-19 Prediction Applying Supervised *Machine learning* Algorithms with Comparative Analysis using WEKA [12]”. Penelitian ini digunakan optimasi menggunakan *hyperparameter* sehingga dapat memaksimalkan nilai akurasi. Hasil penelitian ini menunjukkan bahwa akurasi tertinggi sebesar 98,81% dengan algoritma SVM. Penelitian “*Support Vector Machine* Modelling for COVID-19 Prediction based on Symptoms using R Programming Language [13]” juga melakukan modifikasi dengan menerapkan *hyperparameter* pada algoritma SVM. Akurasi yang diperoleh juga sangat baik yaitu mencapai 98,02%.

Oleh karena itu, apabila dilihat dari penelitian sebelumnya pada penelitian [7-9], penelitian dengan topik *breast cancer* ini hanya menggunakan algoritma SVM saja. Maka penelitian ini akan melakukan modifikasi algoritma SVM dengan menggunakan *kernel* dan menambahkan teknik *hyperparameter*, sebab terdapat peluang untuk dapat meningkatkan akurasi atau menghasilkan akurasi yang lebih tinggi pada penelitian ini. Algoritma SVM sebelumnya juga belum pernah dimodifikasi dengan *hyperparameter* dan diterapkan pada data *breast cancer*, sehingga pada penelitian kali ini akan dilakukan modifikasi pada algoritma *Support Vector Machine* (SVM) menggunakan *Kernel Linear* dan RBF serta dengan menggunakan *hyperparameter*.

## 1.2 Rumusan Masalah

Berdasarkan rumusan latar belakang yang sudah dijelaskan sebelumnya, maka dapat ditarik rumusan masalah. Berikut merupakan beberapa rumusan masalah dalam penelitian ini :

- 1.2.1 Bagaimana prediksi kanker payudara menggunakan algoritma *Support Vector Machine* (SVM) dengan *Kernel Linear*?
- 1.2.2 Bagaimana prediksi kanker payudara menggunakan algoritma *Support Vector Machine* (SVM) dengan *Kernel RBF*?

- 1.2.3 Bagaimana prediksi kanker payudara menggunakan algoritma *Support Vector Machine (SVM)* dengan *hyperparameter*?
- 1.2.4 Bagaimana hasil akhir perbandingan kualitas prediksi ketiga solusi di atas?
- 1.2.5 Apakah model terbaik dari solusi di atas dapat mengalahkan penelitian terdahulu yang menggunakan SVM saja?

### **1.3 Batasan Masalah**

Ada beberapa keterbatasan-keterbatasan dari penelitian ini. Berikut merupakan beberapa keterbatasan masalah penelitian ini :

- 1.3.1 Penelitian ini hanya terbatas pada modifikasi algoritma *Support Vector Machine (SVM)* dengan *Kernel Linear*, RBF, dan *hyperparameter* saja.
- 1.3.2 Penelitian ini hanya berfokus pada topik kanker payudara saja.
- 1.3.3 Penelitian ini hanya sampai pada evaluasi saja, tidak sampai *deployment*.

### **1.4 Tujuan Penelitian dan Manfaat Penelitian**

#### **1.4.1 Tujuan Penelitian**

Dari penelitian yang dilakukan ini bertujuan untuk :

1. Melakukan prediksi dan membandingkan/melakukan komparasi hasil dari *Kernel Linear*, *Kernel RBF*, dan *hyperparameter* dengan menggunakan algoritma *Support Vector Machine*.
2. Dapat memperoleh kombinasi terbaik dari hasil prediksi di atas.

#### **1.4.2 Manfaat Penelitian**

Beberapa manfaat dari penelitian ini yaitu :

1. Dapat membantu dalam bidang kesehatan terutama dalam memprediksi penyakit kanker payudara bahwa deteksi dini itu penting dan perlu dilakukan salah satunya dengan pemodelan *data mining*.
2. Dapat menambah informasi dan pengetahuan baru mengenai pengimplementasian *data mining* terutama dalam menggunakan

algoritma *Support Vector Machine* (SVM), penggunaan *Kernel*, dan *hyperparameter* sehingga dapat membantu dalam dunia akademik.

### 1.5 Sistematika Penulisan

Sistematika penulisan pada penelitian tugas akhir ini antara lain sebagai berikut:

#### BAB I PENDAHULUAN

Pada bagian pendahuluan akan membahas mengenai gambaran umum mengenai prediksi dini kanker payudara yang akan diselesaikan dalam penulisan tugas akhir ini. Biasanya mencakup bagian latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, dan manfaat penelitian dengan menggunakan solusi yang dipilih. Di bagian ini juga akan dijelaskan mengenai sistematika penulisan laporan akhir.

#### BAB II TINJAUAN PUSTAKA

Pada bagian tinjauan pustaka akan dibahas mengenai teori-teori yang digunakan untuk menunjang penelitian tugas akhir. Teori-teori tersebut diantaranya *Support Vector Machine*, *hyperparameter*, CRISP-DM, dan *Confusion matrix*. Kemudian juga akan dibahas mengenai penelitian terdahulu yang menjadi acuan untuk penelitian ini.

#### BAB III METODE PENELITIAN

Pada bagian ini akan dibahas secara rinci mengenai gambaran umum dari objek penelitian, metode penelitian, variable penelitian, teknik pengumpulan data, dan juga teknik analisis data yang digunakan dalam penelitian kanker payudara ini.

#### BAB IV ANALISIS DAN HASIL PENELITIAN



Pada bagian ini akan dilakukan pembahasan lebih mendalam mengenai model yang dibuat yaitu dengan menggunakan algoritma *Support Vector Machine* dengan *Kernel Linear*, *Kernel RBF*, dan *hyperparameter*. Kemudian juga dilakukan analisa lebih lanjut terkait hasil dari pemodelan yang dibuat. Tahapan ini bertujuan untuk menjawab rumusan-rumusan masalah yang ada.

## BAB V

### KESIMPULAN DAN SARAN

Pada bagian ini akan dilakukan pembahasan mengenai kesimpulan yang diperoleh dari pemodelan dan analisis yang dilakukan. Selain itu, akan diberikan saran terkait penelitian yang bertujuan untuk pengembangan penelitian lebih lanjut.

