

KELOMPOK 3

HO-LUNG

HOW ARE YOU LUNG?



TEAM



Robby Data Engineer



Saiful Manager



Fery
Data Scientist



Nabilla Business Development



Dea
Product Development

OUTLINE

Topics for discussion

- Business Understanding
- **2** Tujuan
- **3** Manfaat
- 4 Data Understanding
- 5 Data Preparation
- 6 Modelling
- **7** Evaluation
- 8 Deployment



BUSINESS UNDERSTANDING

HO-LUNG adalah sebuah sistem yang dimanfaatkan dalam bidang kesehatan yang berfokus pada kesehatan paru-paru manusia. Diharapkan dengan menggunakan sistem prediksi yang kami buat pengguna tidak perlu repot melakukan konsultasi tentang gejala yang dialami dengan dokter terlebih dahulu.



TUJUAN



1. Sistem dapat memprediksi kanker paru berdasarkan klasifikasi gejala yang relevan.

2. Dapat memprediksi kanker paru secara efisien dan mudah bagi seluruh kalangan.

MANFAAT

Memprediksi secara dini kanker paruparu dengan mengetahui gejala-gejala yang dialami



Pencegahan dini terhadap bahayanya kanker paru-paru

DATA UNDERSTANDING

- ALLERGY
- ALCOHOL CONSUMING
- SWALLOWING DIFFICULTY
- WHEEZING
- COUGHING
- CHEST PAIN
- PEER_PRESSURE
- YELLOW_FINGERS
- AGE
- ANXIETY
- FATIGUE
- CHRONIC DISEASE
- GENDER
- SHORTNESS OF BREATH
- SMOKING

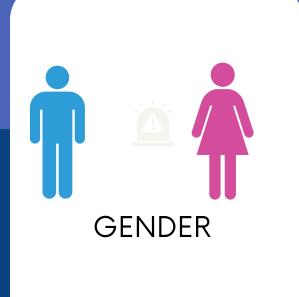








DATA PREPARATION OF LUNG CANCER





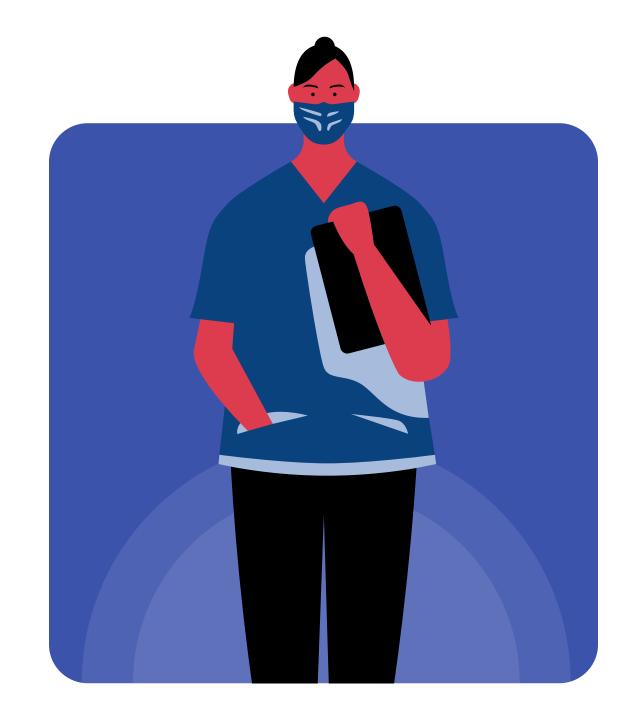
AGE







smoking



MODELLING

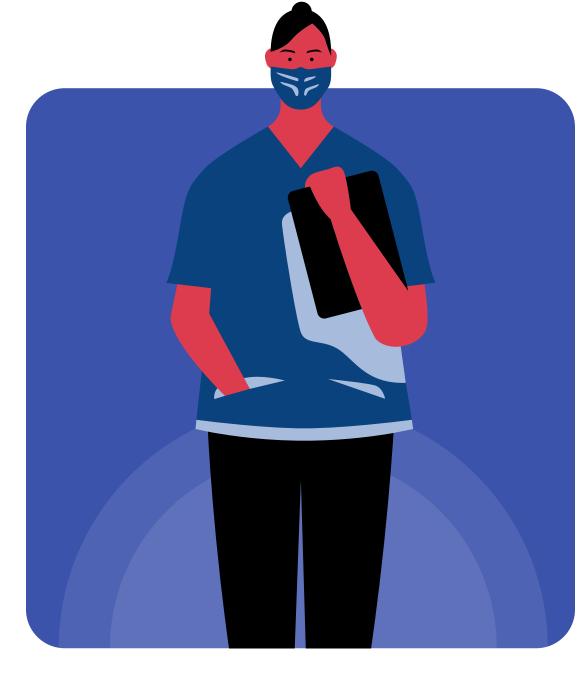
Algoritma yang digunakan adalah Support Vector Classifier dan Logistic Regression

```
#logistic regression with five-fold cross validation
lr = LogisticRegression(max_iter = 2000)
cv = cross_val_score(lr,X_train,y_train,cv=5)
print(mean(cv), '+/-', std(cv))
```

```
#support vector classifier with five-fold cross validation
svc = SVC(probability = True)
cv = cross_val_score(svc,X_train,y_train,cv=5)
print(mean(cv), '+/-', std(cv))
```

0.9351526364477335 + / - 0.023485469326540147

0.9308048103607771 + / - 0.0158848177577188



SVC

print('SVC')
print(classification_report(y_test, y_pred_svc))

<mark>SVC</mark>

support	f1-score	recall	precision		
10	0.56	0.50	0.62	NO	
68	0.94	0.96	0.93	YES	
78	0.90			accuracy	
78	0.75	0.73	0.78	macro avg	
78	0.89	0.90	0.89	weighted avg	



LOGISTIC REGRESSION

```
print('LogisticRegression')
print(classification_report(y_test, y_pred_lr))
```

LogisticRegression

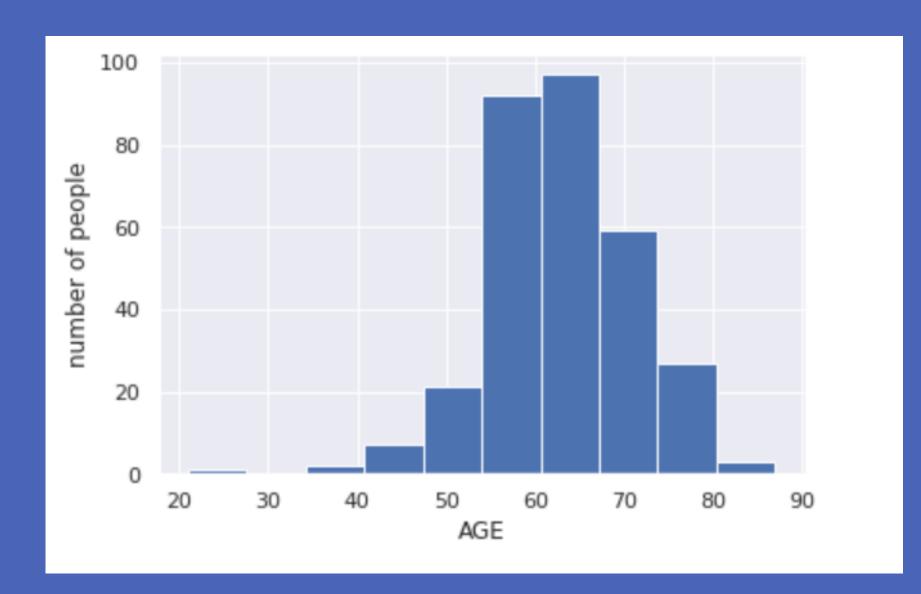
support	f1-score	recall	precision	
10	0.25	0.20	0.33	NO
68	0.91	0.94	0.89	YES
78	0.85			accuracy
78	0.58	0.57	0.61	macro avg
78	0.83	0.85	0.82	weighted avg

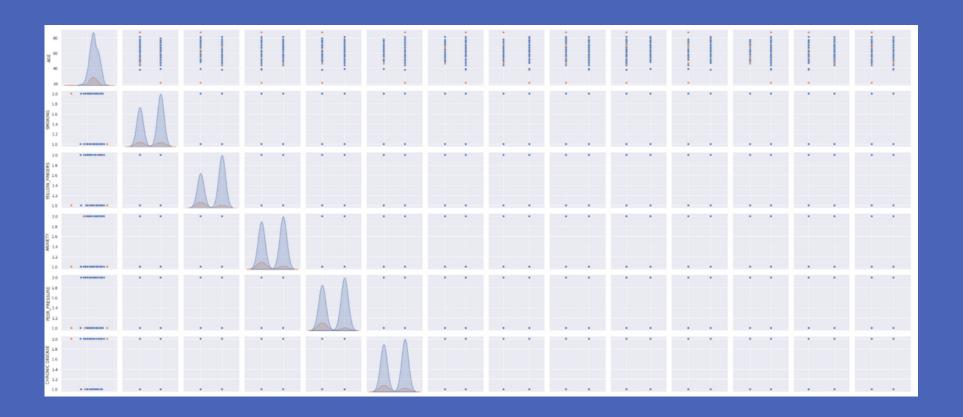
SAVE YOUR LUNG

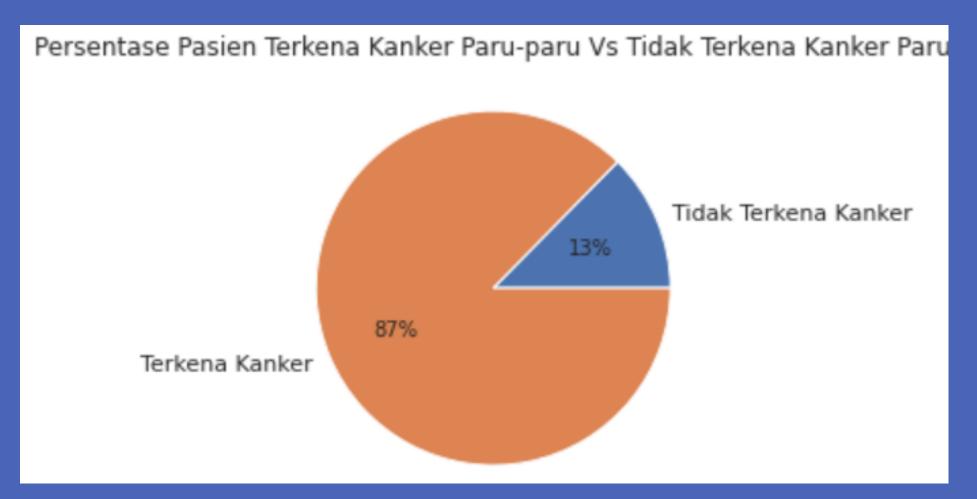
EVALUATION

Dilihat dari dua algoritma yang digunakan mendapatkan Hasil Support Vector Classifier 93% dan Logistic Regression 89%.



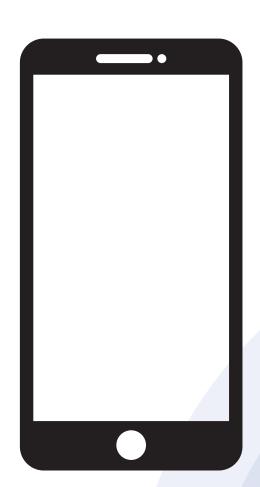






AGE	1	-0.084	0.0052	0.053	0019	4.013	0.013	0 028		0.059	017	4.018	4.0013	4.018			
SMOKING	40.004	1	4.015	016	0.043	-0.14	4.03	0.0019	4.13	0.051	413	0.061	0.031	012			
YELLOW_FINGERS	0.0052	0.015	1	057	0.32	0.041	0.12	0.14	0.079	0.29	0.013	0.11	0.35	01		-08	
ANXIETY	0.053	016	0.57	1	0.22	0.0097	0.19	0.17	4.19	0.17	0.23	-0.14	0.49	411			
PEER_PRESSURE	0.019	0.043	0.32	622	1	0.049	0.078	0.082	0.069	416	0.069	0.22	0.37	4.095		- 0.6	
CHRONIC DISEASE	4013	40.14	0.041	-0 0097	0.049	1	-0.11	0.11	0.05	0.0022	4.18	-0.026	0.075	4.037			
FATIGUE	0.013	40.03	0.12	0.19	0.078	4.11	1	0.0031	0.14	0.19	015	044	4.13	4.011		- 0.4	
ALLERGY	0.028	0.0019	-0.14	417	-0.082	011	0.0031	2	017	0.34	019	4.03	-0.062	0.24			
WHEEZING	0.055	0.13	4.079	0.19	0.069	4.05	014	0.17	1	9.27	0.37	0.938	0.069	0.15		-02	
ALCOHOL CONSUMING	0.059	-0.051	0.29	0.17	0.16	0.0022	0.19	0.54	627	1	62	0.18	-0.0093	633			
COUGHING	017	-0.13	4.013	0.23	-0.089	4.18	015	0.19	0.37	02	1	028	4.16	0.084		-0.0	
SHORTNESS OF BREATH	4.018	0.061	0.11	0.14	4.22	0.026	044	4.03	0.038	418	028	1	4.16	0.024	П		
SWALLOWING DIFFICULTY	0.0013	0.031	0,35		0.37	0.075	4.13	0.062	0.069	4.0093	0.16	0.16	1	0.069			
CHEST PAIN	4.018	012	-01	011	A 095	0.037	-0 011	0.24	015	0.33	0.064	0.024	0.069	1		0.2	
•	MGE	SMOKING	YELLOW_FINGERS	ANKIETY	MEER PRESSURE	CHRONIC DISEASE	FATIGUE	ALLERGY	WHEEZING	ALCOHOL CONSUMING	COUGHING	SHORTNESS OF BALLATH	SWALLOWING DIFFICULTY	CHEST PAIN			





Kemungkinan terbesar model sistem akan diimplementasikan dengan menggunakan website sehingga dapat digunakan secara fleksibel dan dapat diakses oleh semua kalangan. Hingga saat ini kami masih mengembangkan fitur apa saja yang relevan dan dibutuhkan serta design dari user interface maupun user experience sistem ini.

REFERENCE

1. KAGGLE

https://www.kaggle.com/mysarahmadbhat/lung-cancer

2. GOOGLE COLAB

https://colab.research.google.com/drive/1XlyyfEcQ4oMw-

d_MS3FaeTORP7sEVAGV?usp=sharing#scrollTo=m-rQiOSDZNvK

