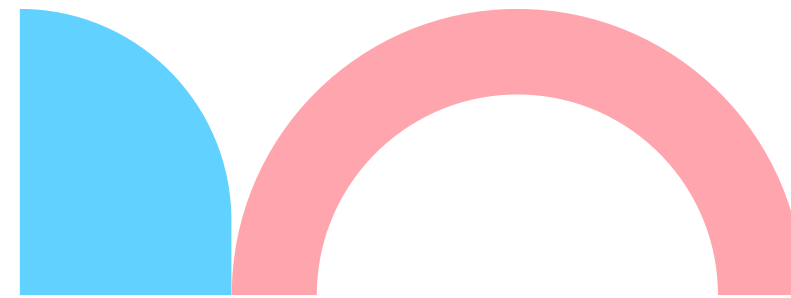


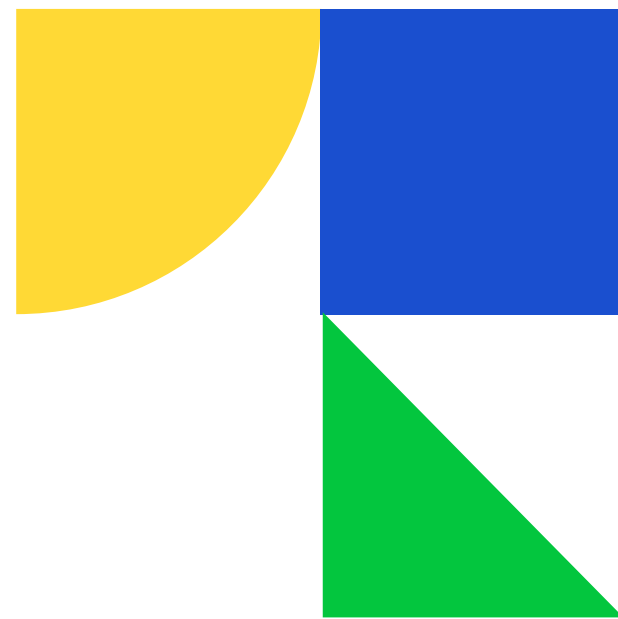
fitria ,Minggu kmarin ketemu
kaamu di jalan,kaamu sdh tidak
kenal aku,coba lihat fotoku
chat-v.com/s/9dtu8y



❖ **KELOMPOK 5**

Deteksi

4jukan segera
(pinjol) dengn mud4h tanpa potongn
Proses Cpat cair
5jt/500jt
info WA langsung
nomor [+6285651275857](https://wa.me/6285651275857)



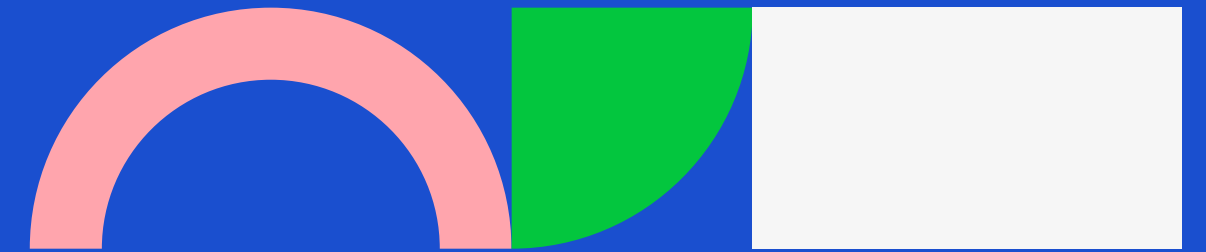
Spam

adit ,kemarin reunian
sama guru kita kamu ngak
datang,inli yang sa_ya f0to
chat-v.com/s/9jp19n

Elvano Jethro Mogi Pardede - 2502025403
Rasyad Muhammad Ramdhanazuri - 2501992305
Theophilus Nathan Tanudjaja - 2502007514
Wicakra Hanly Tanudjaja - 2502032964

[View Slideshow](#)

Latar Belakang



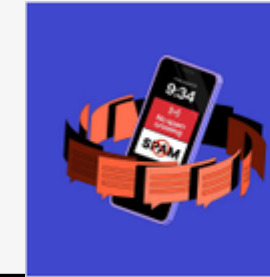
Saat ini, banyak sekali SMS spam yang mengganggu pengguna. Menurut laporan dari Kementerian Komunikasi dan Informatika (Kominfo), selama Agustus 2018 hingga 16 Februari 2023, terdapat 1.730 konten penipuan online di Indonesia. Kerugian akibat penipuan online mencapai Rp 18,7 triliun selama periode 2017 hingga 2021. Selain itu, rata-rata orang Indonesia menerima 14 panggilan spam setiap hari, dan sekitar 66,6% pernah menjadi korban penipuan online. Oleh karena itu, deteksi SMS spam menjadi sangat penting.

Dengan melakukan deteksi spam, kita dapat:

- **Melindungi pengguna**
- **Efisien dalam komunikasi**
- **Menghemat waktu dan energi**
- **Mencegah penipuan**

Dengan demikian, deteksi SMS spam adalah langkah penting untuk melindungi pengguna dan meningkatkan kualitas komunikasi mereka.

Dataset



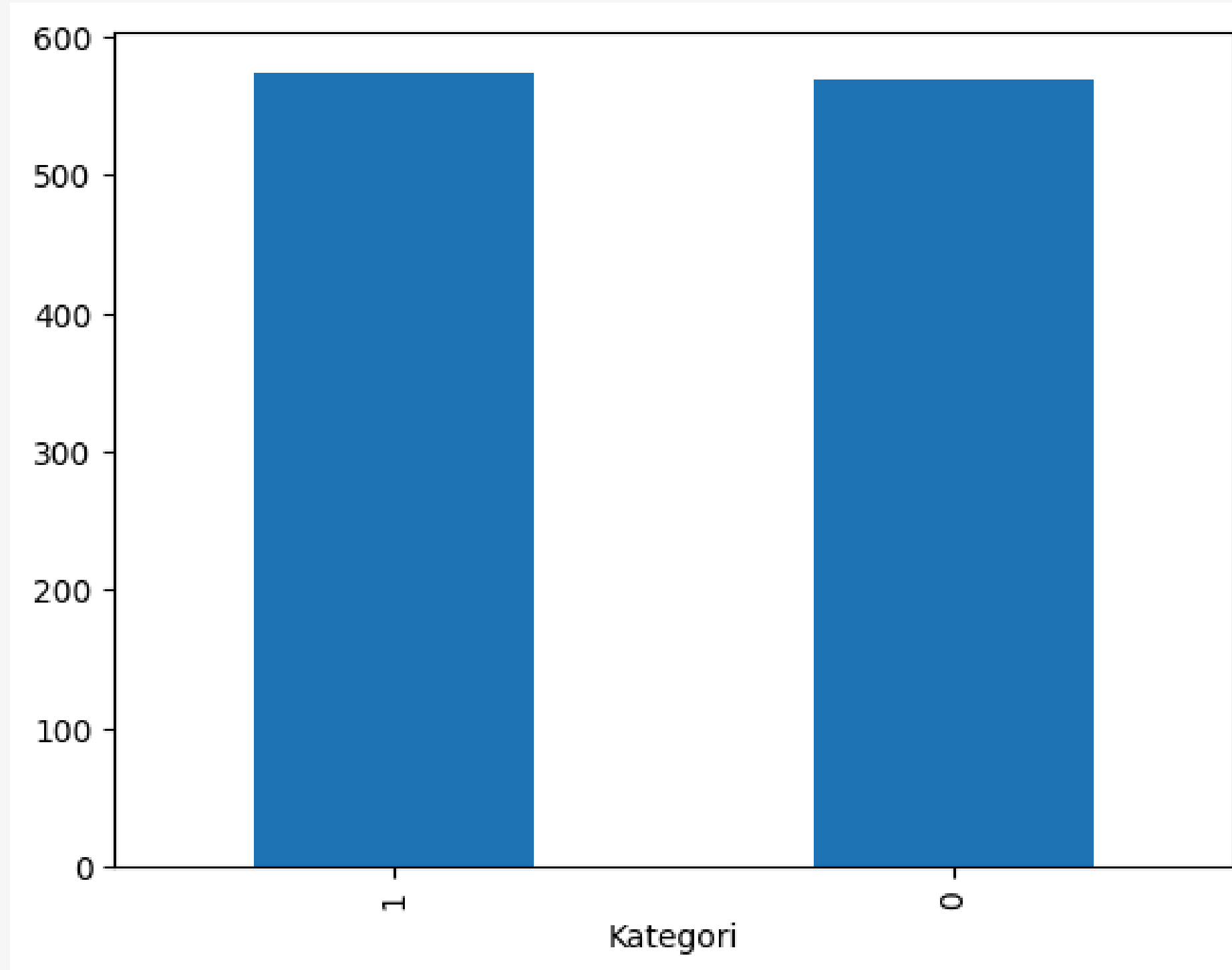
Indonesian SMS Spam

Dataset of Spam SMS (Messages) in Indonesian Language

[kaggle.com](https://www.kaggle.com)

| Kategori | Pesan |
|----------|--|
| spam | <i>Plg Yth: Simcard anda mendptkan bonus poin plus-plus 555 dr:PT.INDOSAT pin anda:277fg49 u/info klik di www.indosat-555.blogspot.com atau Hub:021-3338-0074.</i> |
| ham | <i>Iya ih ko sedih sih gtau kapan lg ke bandung :(</i> |
| ham | <i>Kalau mau bikin model/controller mending per apa y?</i> |
| ham | <i>Selamat nama1. Semoga selalu menempuh hidup yang bahagia dan selalu menyenangkan</i> |
| spam | <i>Tingkatkan nilai isi ulang Anda selanjutnya minimal Rp10ribu untuk bisa mendapatkan Paket MURAH PAS di *100*999#. Info 188.</i> |
| spam | <i>Uangnya di Trnsfer di BNI ini aja a/n RASTINA NO.REK.0253 7915 37.Klo sdh d Trnsfer Sms aja di no ini 085293355598 Trmks.</i> |
| spam | <i>PESAN RESMI Selamat NO Simcard anda M-dptkam hadiah cek.Rp.35juta dr TRI CARE pin:7Y7R8K9Z info keik: www.id-tricare2016.blogspot.com</i> |
| spam | <i>Sy RUDI di jambi, bgi tman2 yg sering kalah Togel saatnya kalahkan bandar hub mbah ROS 082325557768 di jamin tembus, sy sdh buktikan 4D 50jt.</i> |
| | <i>Tri berBAGI bonus PULSA di bulan penuh berkah! Isi pulsa min. 5rb dapet</i> |

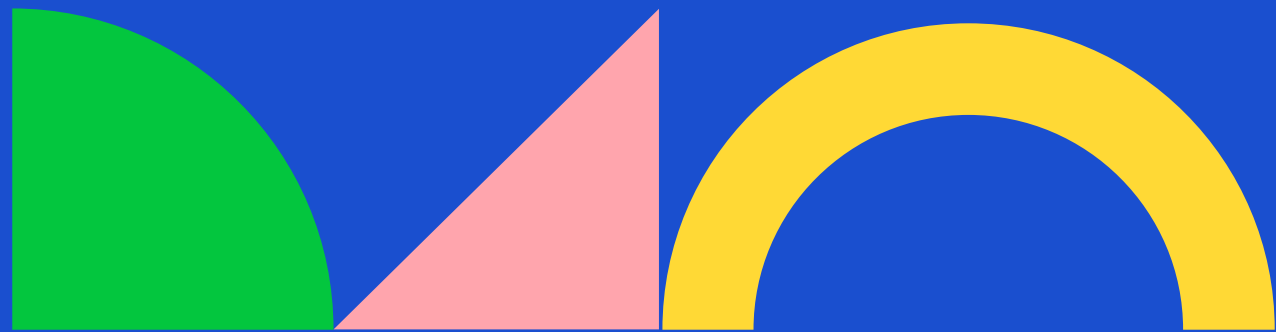
Distribusi Label



Dimana
1: spam
0: ham



Data Preprocessing



Jik;2A@jik;2a

Case Folding

jik;2a@jika

Noise Removal

tingkatkan@tingkat

Stemming

tingkat nilai
@
[tingkat, nilai]

Tokenizing

[putri, dan, adison]
@
[putri, adison]

Hapus stop words

[putri, adison]
@
putri adison

Gabung kembali jadi String

Case Folding

Jik;2A©jik;2a

| Raw | Case Folding |
|--|--|
| <i>Plg Yth: Simcard anda mendptkan bonus poin plus-plus 555 dr:PT.INDOSAT pin anda:277fg49 u/info klik di www.indosat-555.blogspot.com atau Hub:021-3338-0074.</i> | <i>plg yth: simcard anda mendptkan bonus poin plus-plus 555 dr:pt.indosat pin anda:277fg49 u/info klik di www.indosat-555.blogspot.com atau hub:021-3338-0074.</i> |
| <i>Iya ih ko sedih sih gtau kapan lg ke bandung :(</i> | <i>iya ih ko sedih sih gtau kapan lg ke bandung :(</i> |
| <i>Kalau mau bikin model/controller mending per apa y?</i> | <i>kalau mau bikin model/controller mending per apa y?</i> |
| <i>Selamat nama1. Semoga selalu menempuh hidup yang bahagia dan selalu menyenangkan</i> | <i>selamat nama1. semoga selalu menempuh hidup yang bahagia dan selalu menyenangkan</i> |
| <i>Tingkatkan nilai isi ulang Anda selanjutnya minimal Rp10ribu untuk bisa mendapatkan Paket MURAH PAS di *100*999#. Info 188.</i> | <i>tingkatkan nilai isi ulang anda selanjutnya minimal rp10ribu untuk bisa mendapatkan paket murah pas di *100*999#. info 188.</i> |
| <i>Uangnya di Trnsfer di BNI ini aja a/n RASTINA NO.REK.0253 7915 37.Klo sdh d Trnsfer Sms aja di no ini 085293355598 Trmks.</i> | <i>uangnya di trnsfer di bni ini aja a/n rastina no.rek.0253 7915 37.klo sdh d trnsfer sms aja di no ini 085293355598 trmks.</i> |
| <i>PESAN RESMI Selamat NO Simcard anda M-dptkam hadiah cek.Rp.35jutadr TRI CARE pin:7Y7R8K9Z info keik: www.id-tricore2016.blogspot.com</i> | <i>pesan resmi selamat no simcard anda m-dptkam hadiah cek.rp.35juta dr tri care pin:7y7r8k9z info keik: www.id-tricore2016.blogspot.com</i> |

Noise Removal

jik;2a@jika

| Case Folding | Noise Removal |
|--|---|
| <i>plg yth: simcard anda mendptkan bonus poin plus-plus 555 dr:pt.indosat pin anda:277fg49 u/info klik di www.indosat-555.blogspot.com atau hub:021-3338-0074.</i> | <i>plg yth simcard anda mendptkan bonus poin plus plus dr pt indosat pin anda fg u info klik di w indosat blogspot com atau hub</i> |
| <i>iya ih ko sedih sih gtau kapan lg ke bandung :(</i> | <i>iya ih ko sedih sih gtau kapan lg ke bandung</i> |
| <i>kalau mau bikin model/controller mending per apa y?</i> | <i>kalau mau bikin model controller mending per apa y</i> |
| <i>selamat nama1. semoga selalu menempuh hidup yang bahagia dan selalu menyenangkan</i> | <i>selamat nama semoga selalu menempuh hidup yang bahagia dan selalu menyenangkan</i> |
| <i>tingkatkan nilai isi ulang anda selanjutnya minimal rp10ribu untuk bisa mendapatkan paket murah pas di *100*999#. info 188.</i> | <i>tingkatkan nilai isi ulang anda selanjutnya minimal rpribu untuk bisa mendapatkan paket murah pas di info</i> |
| <i>uangnya di trnsfer di bni ini aja a/n rastina no.rek.0253 7915 37.klo sdh d trnsfer sms aja di no ini 085293355598 trmks.</i> | <i>uangnya di trnsfer di bni ini aja a n rastina no rek klo sdh d trnsfer sms aja di no ini trmks</i> |
| <i>pesan resmi selamat no simcard anda m-dptkam hadiah cek.rp.35juta dr tri care pin:7y7r8k9z info keik: www.id-tricare2016.blogspot.com</i> | <i>pesan resmi selamat no simcard anda m dptkam hadiah cek rp juta dr tri care pin yrkz info keik w id tricare blogspot com</i> |

Stemming

tingkatkan@tingkat

| Noise Removal | Stemming |
|---|---|
| <i>plg yth simcard anda mendptkan bonus poin plus plus dr pt indosat pin anda fg u info klik di w indosat blogspot com atau hub</i> | <i>plg yth simcard anda mendptkan bonus poin plus plus dr pt indosat pin anda fg u info klik di w indosat blogspot com atau hub</i> |
| <i>iya ih ko sedih sih gtau kapan lg ke bandung</i> | <i>iya ih ko sedih sih gtau kapan lg ke bandung</i> |
| <i>kalau mau bikin model controller mending per apa y</i> | <i>kalau mau bikin model controller mending per apa y</i> |
| <i>selamat nama semoga selalu menempuh hidup yang bahagia dan selalu menyenangkan</i> | <i>selamat nama moga selalu tempuh hidup yang bahagia dan selalu senang</i> |
| <i>tingkatkan nilai isi ulang anda selanjutnya minimal rpribu untuk bisa mendapatkan paket murah pas di info</i> | <i>tingkat nilai isi ulang anda lanjut minimal rpribu untuk bisa dapat paket murah pas di info</i> |
| <i>uangnya di trnsfer di bni ini aja a n rastina no rek klo sdh d trnsfer sms aja di no ini trmks</i> | <i>uang di trnsfer di bni ini aja a n rastina no rek klo sdh d trnsfer sms aja di no ini trmks</i> |
| <i>pesan resmi selamat no simcard anda m dptkam hadiah cek rp juta dr tri care pin yrkz info keik w id tricare blogspot com</i> | <i>pesan resmi selamat no simcard anda m dptkam hadiah cek rp juta dr tri care pin yrkz info keik w id tricare blogspot com</i> |

Tokenizing

tingkat nilai@[tingkat, nilai]

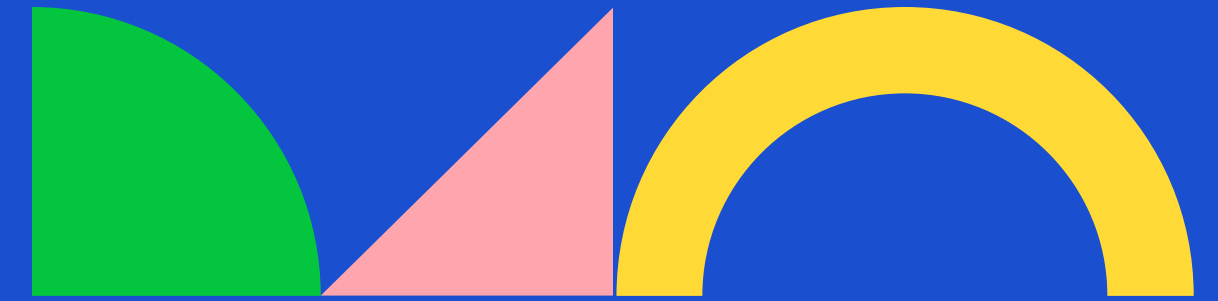
| Stemming | Tokenizing |
|---|---|
| <i>plg yth simcard anda mendptkan bonus poin plus plus dr pt indosat pin anda fg u info klik di w indosat blogspot com atau hub</i> | <i>['plg', 'yth', 'simcard', 'anda', 'mendptkan', 'bonus', 'poin', 'plus', 'plus', 'dr', 'pt', 'indosat', 'pin', 'anda', 'fg', 'u', 'info', 'klik', 'di', 'w', 'indosat', 'blogspot', 'com', 'atau', 'hub']</i> |
| <i>iya ih ko sedih sih gtau kapan lg ke bandung</i> | <i>['iya', 'ih', 'ko', 'sedih', 'sih', 'gtau', 'kapan', 'lg', 'ke', 'bandung']</i> |
| <i>kalau mau bikin model controller mending per apa y</i> | <i>['kalau', 'mau', 'bikin', 'model', 'controller', 'mending', 'per', 'apa', 'y']</i> |
| <i>selamat nama moga selalu tempuh hidup yang bahagia dan selalu senang</i> | <i>['selamat', 'nama', 'moga', 'selalu', 'tempuh', 'hidup', 'yang', 'bahagia', 'dan', 'selalu', 'senang']</i> |
| <i>tingkat nilai isi ulang anda lanjut minimal rpribu untuk bisa dapat paket murah pas di info</i> | <i>['tingkat', 'nilai', 'isi', 'ulang', 'anda', 'lanjut', 'minimal', 'rpribu', 'untuk', 'bisa', 'dapat', 'paket', 'murah', 'pas', 'di', 'info']</i> |
| <i>uang di trnsfer di bni ini aja a n rastina no rek klo sdh d trnsfer sms aja di no ini trmks</i> | <i>['uang', 'di', 'trnsfer', 'di', 'bni', 'ini', 'aja', 'a', 'n', 'rastina', 'no', 'rek', 'klo', 'sdh', 'd', 'trnsfer', 'sms', 'aja', 'di', 'no', 'ini', 'trmks']</i> |
| <i>pesan resmi selamat no simcard anda m dptkam hadiah cek rp juta dr tri care pin yrkz info keik w id tricare blogspot com</i> | <i>['pesan', 'resmi', 'selamat', 'no', 'simcard', 'anda', 'm', 'dptkam', 'hadiah', 'cek', 'rp', 'juta', 'dr', 'tri', 'care', 'pin', 'yrkz', 'info', 'keik', 'w', 'id', 'tricare', 'blogspot', 'com']</i> |

Hapus Stop Word

[putri, dan, adison] → [putri, adison]

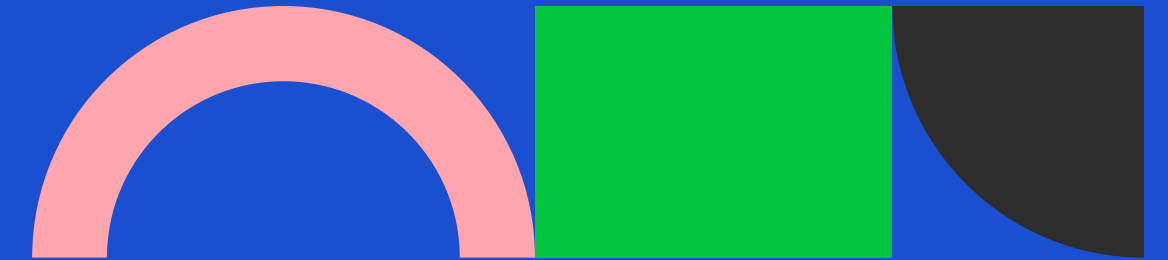
| Tokenizing | Stemming | Lalu digabungkan kembali menjadi string |
|---|--|---|
| <i>['plg', 'yth', 'simcard', 'anda', 'mendptkan', 'bonus', 'poin', 'plus', 'plus', 'dr', 'pt', 'indosat', 'pin', 'anda', 'fg', 'u', 'info', 'klik', 'di', 'w', 'indosat', 'blogspot', 'com', 'atau', 'hub']</i> | <i>['plg', 'yth', 'simcard', 'mendptkan', 'bonus', 'poin', 'plus', 'plus', 'dr', 'pt', 'indosat', 'pin', 'fg', 'u', 'info', 'klik', 'w', 'indosat', 'blogspot', 'com', 'hub']</i> | |
| <i>['iya', 'ih', 'ko', 'sedih', 'sih', 'gtau', 'kapan', 'lg', 'ke', 'bandung']</i> | <i>['iya', 'ih', 'ko', 'sedih', 'sih', 'gtau', 'lg', 'bandung']</i> | |
| <i>['kalau', 'mau', 'bikin', 'model', 'controller', 'mending', 'per', 'apa', 'y']</i> | <i>['bikin', 'model', 'controller', 'mending', 'y']</i> | |
| <i>['selamat', 'nama', 'moga', 'selalu', 'tempuh', 'hidup', 'yang', 'bahagia', 'dan', 'selalu', 'senang']</i> | <i>['selamat', 'nama', 'moga', 'tempuh', 'hidup', 'bahagia', 'senang']</i> | |
| <i>['tingkat', 'nilai', 'isi', 'ulang', 'anda', 'lanjut', 'minimal', 'rpribu', 'untuk', 'bisa', 'dapat', 'paket', 'murah', 'pas', 'di', 'info']</i> | <i>['tingkat', 'nilai', 'isi', 'ulang', 'minimal', 'rpribu', 'paket', 'murah', 'pas', 'info']</i> | |
| <i>['uang', 'di', 'trnsfer', 'di', 'bni', 'ini', 'aja', 'a', 'n', 'rastina', 'no', 'rek', 'klo', 'sdh', 'd', 'trnsfer', 'sms', 'aja', 'di', 'no', 'ini', 'trmks']</i> | <i>['uang', 'trnsfer', 'bni', 'aja', 'a', 'n', 'rastina', 'no', 'rek', 'klo', 'sdh', 'd', 'trnsfer', 'sms', 'aja', 'no', 'trmks']</i> | |
| <i>['pesan', 'resmi', 'selamat', 'no', 'simcard', 'anda', 'm', 'dptkam', 'hadiah', 'cek', 'rp', 'juta', 'dr', 'tri', 'care', 'pin', 'yrkz', 'info', 'keik', 'w', 'id', 'tricare', 'blogspot', 'com']</i> | <i>['pesan', 'resmi', 'selamat', 'no', 'simcard', 'm', 'dptkam', 'hadiah', 'cek', 'rp', 'juta', 'dr', 'tri', 'care', 'pin', 'yrkz', 'info', 'keik', 'w', 'id', 'tricare', 'blogspot', 'com']</i> | |

Cleaned Data



| RAW | CLEANED |
|--|--|
| <i>Plg Yth: Simcard anda mendptkan bonus poin plus-plus 555 dr:PT.INDOSAT pin anda:277fg49 u/info klik di www.indosat-555.blogspot.com atau Hub:021-3338-0074.</i> | <i>plg yth simcard mendptkan bonus poin plus plus dr pt indosat pin fg u info klik w indosat blogspot com hub</i> |
| <i>Iya ih ko sedih sih gtau kapan lg ke bandung :(</i> | <i>iya ih ko sedih sih gtau lg bandung</i> |
| <i>Kalau mau bikin model/controller mending per apa y?</i> | <i>bikin model controller mending y</i> |
| <i>Selamat nama1. Semoga selalu menempuh hidup yang bahagia dan selalu menyenangkan</i> | <i>selamat nama moga tempuh hidup bahagia senang</i> |
| <i>Tingkatkan nilai isi ulang Anda selanjutnya minimal Rp10ribu untuk bisa mendapatkan Paket MURAH PAS di *100*999#. Info 188.</i> | <i>tingkat nilai isi ulang minimal rpribu paket murah pas info</i> |
| <i>Uangnya di Trnsfer di BNI ini aja a/n RASTINA NO.REK.0253 7915 37.Klo sdh d Trnsfer Sms aja di no ini 085293355598 Trmks.</i> | <i>uang trnsfer bni aja a n rastina no rek klo sdh d trnsfer sms aja no trmks</i> |
| <i>PESAN RESMI Selamat NO Simcard anda M-dptkam hadiah cek.Rp.35juta dr TRI CARE pin:7Y7R8K9Z info keik: www.id-tricare2016.blogspot.com</i> | <i>pesan resmi selamat no simcard m dptkam hadiah cek rp juta dr tri care pin yrkz info keik w id tricare blogspot com</i> |

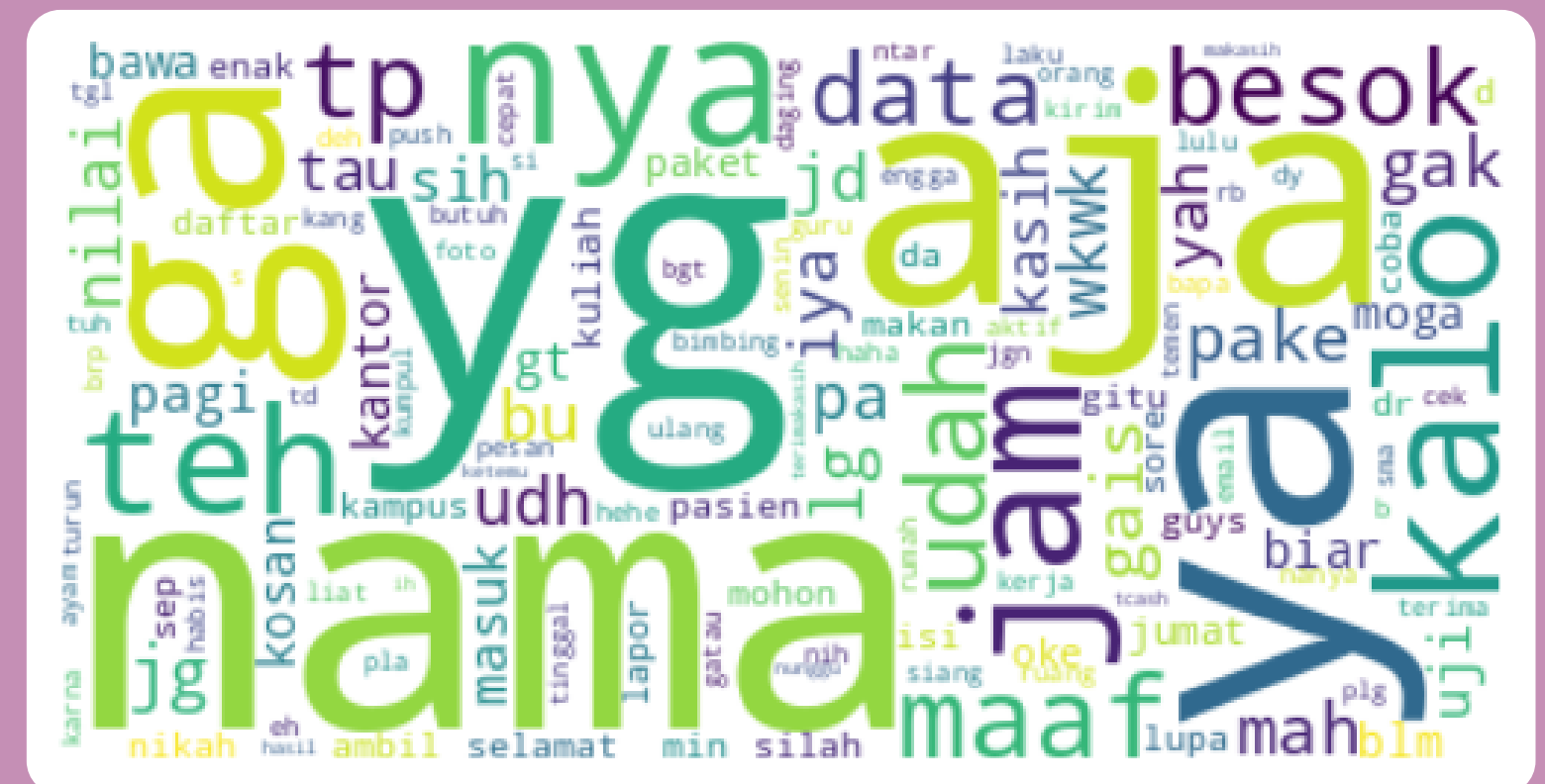
WordCloud



spam



ham



[illegible][illegible]

Melakukan Pemodelan

Membuat Model Indo-BERT

Import
Pre-trained
BERT Model

Tokenize Data
dan
Train Test Split



Buat Keras
Model dan
lakukan
fitting

Membuat Model SVM

Train Test
Split



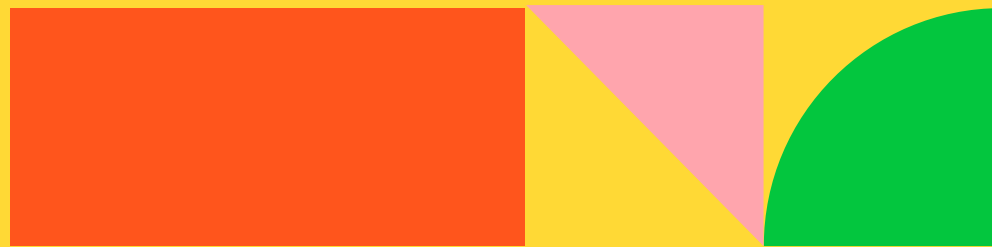
Melakukan text
representation
dengan TF-IDF

Buat Model
SVM dan
lakukan
fitting

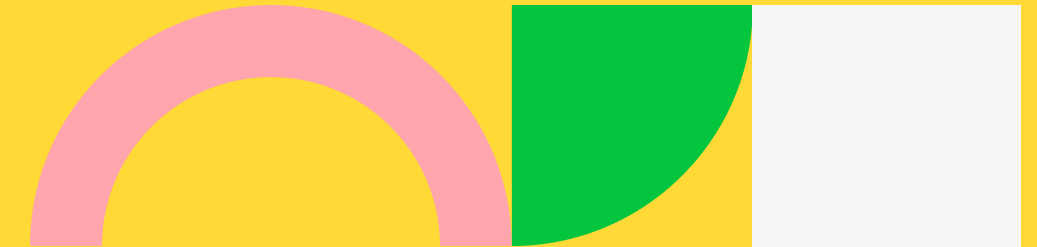
Data
Preprocessing

Indo-BERT

Import pre-trained model



cahya/bert-base-indonesian-522M



- Model ini dilatih menggunakan 522 juta parameter yang di train dari Wikipedia berbahasa Indonesia.
- Teks-teks tersebut telah diubah menjadi huruf kecil dan ditokenisasi menggunakan metode WordPiece dengan ukuran kosakata sebanyak 32.000
- Input dari model ini memiliki bentuk sebagai berikut:
[CLS] Kalimat A [SEP] Kalimat B [SEP]

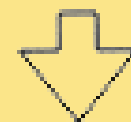
Indo-BERT

Special Tokenization dan Train Test Split

- Special Tokenization

"This is a input text."

Tokenization



| [CLS] | This | is | a | input | . | [SEP] |
|-------|------|------|------|-------|------|-------|
| 101 | 2023 | 2003 | 1037 | 7953 | 1012 | 102 |

**Test
30%**

**Train
70%**

Indo-BERT

Buat Model Keras

✓ Creating a Keras Model

```
[33] max_len = 32

inps = Input(shape = (max_len,), dtype = 'int64')
masks = Input(shape = (max_len,), dtype = 'int64')
indo_bert_layer = bert_model(inps, attention_mask = masks)[0][:, 0, :]
dropout = Dropout(0.5)(indo_bert_layer)
pred = Dense(1, activation = 'sigmoid', kernel_regularizer = regularizers.l2(0.01))(dropout)
model = tf.keras.Model(inputs = [inps, masks], outputs = pred)
```

✓ Compiling the Model

```
[34] losses = BinaryCrossentropy(from_logits = False)
metric = BinaryAccuracy('accuracy')
optimizer = tf.keras.optimizers.AdamW(learning_rate = 5e-5)

model.compile(loss = losses, optimizer = optimizer, metrics = [metric])
```

Indo-BERT

Fitting dan Evaluation Matrix

```
[35] history = model.fit([train_inp, train_mask], train_label, batch_size = 64, epochs = 5)
```

Epoch 1/5

WARNING:tensorflow:Gradients do not exist for variables ['tf_bert_model/bert/pooler/dense/']
WARNING:tensorflow:Gradients do not exist for variables ['tf_bert_model/bert/pooler/dense/']
WARNING:tensorflow:Gradients do not exist for variables ['tf_bert_model/bert/pooler/dense/']
WARNING:tensorflow:Gradients do not exist for variables ['tf_bert_model/bert/pooler/dense/']

13/13 [=====] - 362s 24s/step - loss: 0.9182 - accuracy: 0.6075

Epoch 2/5

13/13 [=====] - 293s 22s/step - loss: 0.4407 - accuracy: 0.8462

Epoch 3/5

13/13 [=====] - 318s 25s/step - loss: 0.1994 - accuracy: 0.9325

Epoch 4/5

13/13 [=====] - 354s 27s/step - loss: 0.1512 - accuracy: 0.9538

Epoch 5/5

13/13 [=====] - 310s 24s/step - loss: 0.0854 - accuracy: 0.9825

- Test Evaluation Matrix

```
11/11 [=====] - 47s 4s/step
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| ham | 0.9806 | 0.9157 | 0.9470 | 166 |
| spam | 0.9255 | 0.9831 | 0.9534 | 177 |
| accuracy | | | 0.9504 | 343 |
| macro avg | 0.9531 | 0.9494 | 0.9502 | 343 |
| weighted avg | 0.9522 | 0.9504 | 0.9503 | 343 |

Indo-BERT

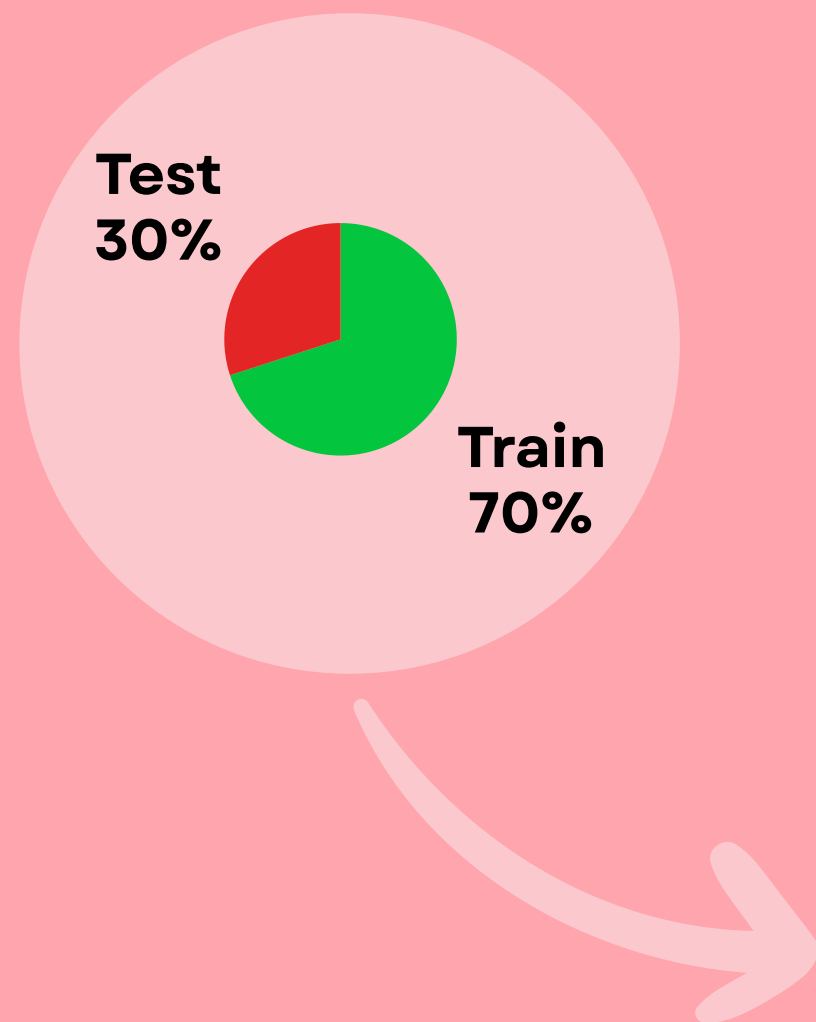
Melakukan Prediksi

```
[ ] def single_predict_bert(sentence):  
    sentence = clean(sentence)  
  
    input_dict = tokenizer.encode_plus(sentence, add_special_tokens = True, max_length = 32, pad_to_max_length = True, return_attention_mask = True, truncation = True)  
    input_ids = tf.constant(input_dict['input_ids'])[None, :]  
    attention_mask = tf.constant(input_dict['attention_mask'])[None, :]  
  
    logits = model.predict([input_ids, attention_mask], verbose = 0)  
    pred = tf.math.argmax(logits, axis = 1).numpy()[0]  
    label_map = {0: 'ham', 1: 'spam'}  
    predicted_label = label_map[pred]  
  
    print(f"Bert Predicted label for the sentence '{sentence}' is: {predicted_label}")
```

```
[ ] single_predict_bert("Aku sayang kamu")  
single_predict_bert("Selamat nomor sim card anda memenangkan hadiah motor beat")
```

```
⇒ Bert Predicted label for the sentence 'Aku sayang kamu' is: ham  
Bert Predicted label for the sentence 'Selamat nomor sim card anda memenangkan hadiah motor beat' is: spam
```

SVM



Text Representation: TF-IDF

Term Frequency - Inverse Document Frequency

- TF: mengukur frekuensi kemunculan sebuah kata dalam sebuah dokumen.
- IDF: mengukur seberapa umum atau jarang suatu kata dalam keseluruhan kumpulan dokumen.
- TF-IDF: $TF \times IDF$

Misal:

A screenshot of a Jupyter Notebook cell showing the output of the code `tfidf_train['info']`. The output is a Series of 800 float64 values. The first five values are 0.000000, 0.000000, 0.148979, 0.000000, and 0.000000. This is followed by an ellipsis (...) and then values for indices 795 through 799, all of which are 0.000000. The bottom of the output shows 'Name: info, Length: 800, dtype: float64'.

| | tfidf_train['info'] |
|---|---------------------|
| 0 | 0.000000 |
| 1 | 0.000000 |
| 2 | 0.148979 |
| 3 | 0.000000 |
| 4 | 0.000000 |
| ... | ... |
| 795 | 0.000000 |
| 796 | 0.000000 |
| 797 | 0.000000 |
| 798 | 0.000000 |
| 799 | 0.000000 |
| Name: info, Length: 800, dtype: float64 | |

SVM

Buat Model SVM



Classificaiton Report Training

```

  Create SVM Model

[40] clf = svm.SVC(kernel = 'sigmoid', random_state = 42)

  Fitting the Model

[41] clf.fit(tfidf_train, df_train['Kategori'])

```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| ham | 1.0000 | 0.9824 | 0.9911 | 398 |
| spam | 0.9829 | 1.0000 | 0.9914 | 402 |
| accuracy | | | 0.9912 | 800 |
| macro avg | 0.9914 | 0.9912 | 0.9912 | 800 |
| weighted avg | 0.9914 | 0.9912 | 0.9912 | 800 |



SVM

Buat Model SVM

Classificaiton Report Testing



| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| ham | 0.9760 | 0.9532 | 0.9645 | 171 |
| spam | 0.9545 | 0.9767 | 0.9655 | 172 |
| accuracy | | | 0.9650 | 343 |
| macro avg | 0.9653 | 0.9650 | 0.9650 | 343 |
| weighted avg | 0.9653 | 0.9650 | 0.9650 | 343 |

SVM

Melakukan Prediksi

```
[ ] def single_predict_svm(sentence):  
    sentence = clean(sentence)  
  
    tfidf_input = vectorizer.transform([sentence])  
    tfidf_input_df = pd.DataFrame(tfidf_input.toarray(), columns = vectorizer.get_feature_names_out())  
  
    # Make prediction  
    predicted_category = clf.predict(tfidf_input_df)[0] # Assuming clf is trained and available  
  
    # Map predicted label to its corresponding category  
    predicted_label = label_map[predicted_category]  
    print(f"SVM Predicted label for the sentence '{sentence}' is: {predicted_label}")
```

```
[ ] single_predict_svm("aku cinta kamu")  
single_predict_svm("Selamat nomor sim card anda memenangkan hadiah motor beat")
```

```
⇒ SVM Predicted label for the sentence 'aku cinta kamu' is: ham  
SVM Predicted label for the sentence 'Selamat nomor sim card anda memenangkan hadiah motor beat' is: spam
```

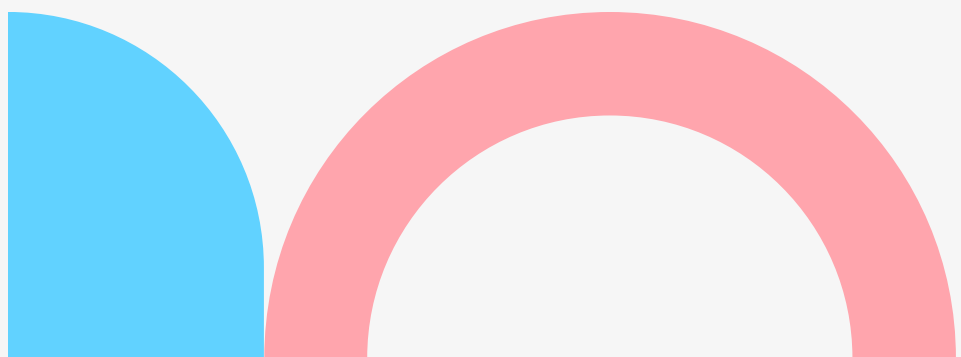
Perbandingan Model

Indo-BERT

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| ham | 0.9806 | 0.9157 | 0.9470 | 166 |
| spam | 0.9255 | 0.9831 | 0.9534 | 177 |
| accuracy | | | 0.9504 | 343 |
| macro avg | 0.9531 | 0.9494 | 0.9502 | 343 |
| weighted avg | 0.9522 | 0.9504 | 0.9503 | 343 |

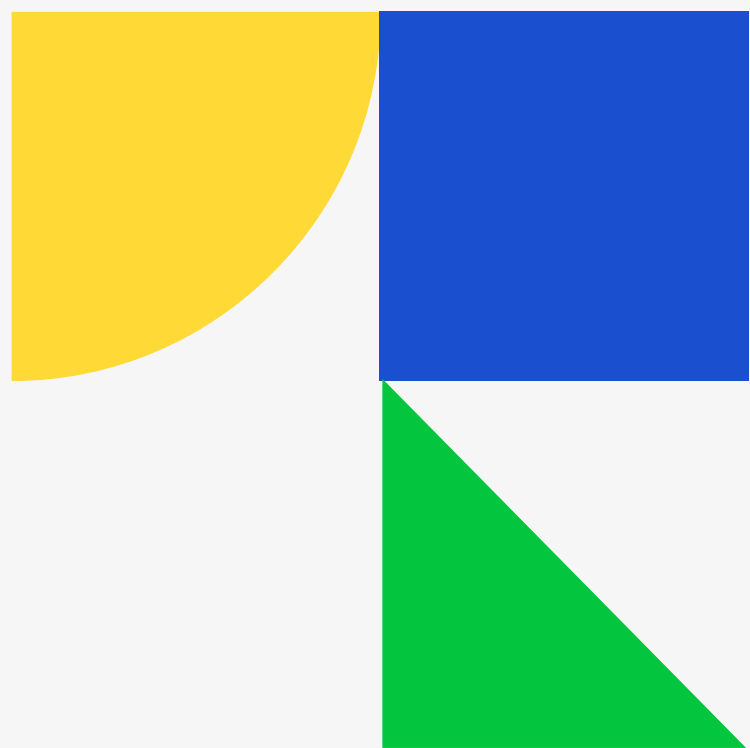
SVM

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| ham | 0.9760 | 0.9532 | 0.9645 | 171 |
| spam | 0.9545 | 0.9767 | 0.9655 | 172 |
| accuracy | | | 0.9650 | 343 |
| macro avg | 0.9653 | 0.9650 | 0.9650 | 343 |
| weighted avg | 0.9653 | 0.9650 | 0.9650 | 343 |



❖ KELOMPOK 5

Terima



Kasih

