

## Daftar Isi

Daftar Isi .....	i
Daftar Tabel & Gambar .....	ii
A. Tentang Aplikasi .....	1
B. Tampilan Aplikasi .....	1
C. Panduan Penggunaan Aplikasi .....	4
D. Detail Proses Aplikasi .....	5
1. Data Preparation .....	5
1.1. <i>Library</i> Pendukung .....	5
1.2. <i>Read Data</i> .....	6
2. Translate Text .....	6
2.1. <i>Typo Text</i> .....	7
2.2. <i>Translate Text</i> .....	8
2.3. <i>Lower</i> .....	10
3. Keyword Extraction .....	11
4. Kategorisasi .....	12
4.1. Persiapan Data .....	12
4.2. <i>Categorization</i> .....	12
4.3. <i>Set Stability</i> .....	14
4.4. Finalisasi Kategori .....	14
5. Name Entity Recognition .....	15
6. <i>Post Processing</i> .....	16

## Daftar Tabel & Gambar

### Tabel

<b>Tabel 1.</b> Komponen dalam Aplikasi.....	2
--	---

### Gambar

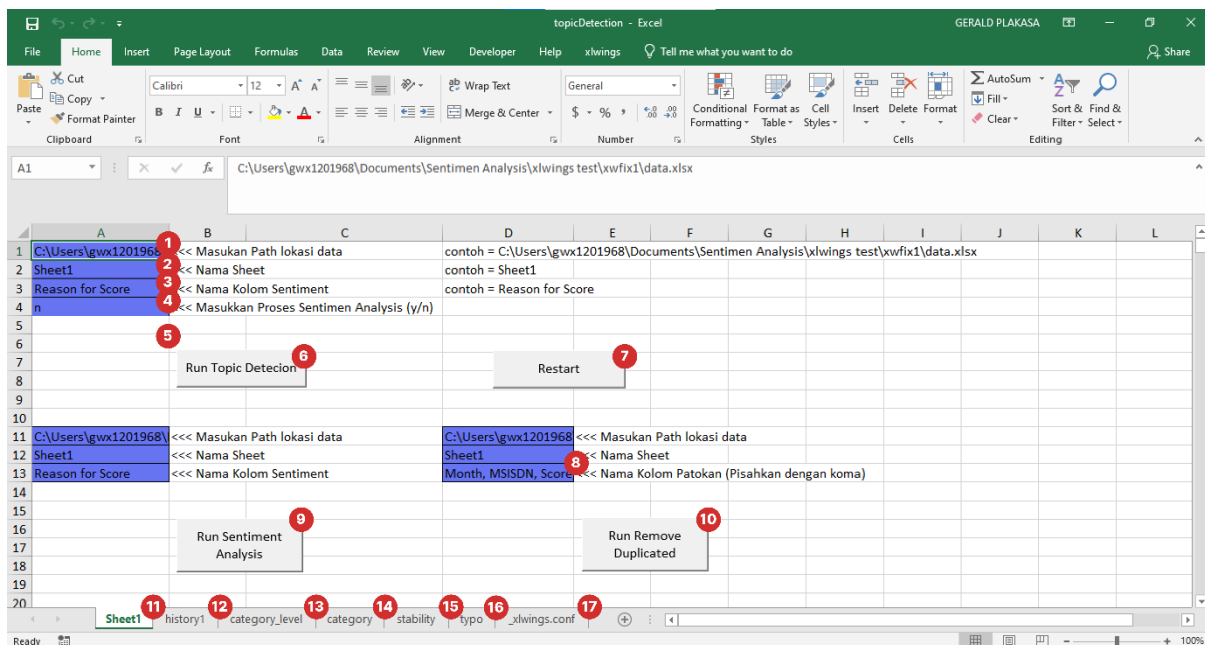
<b>Gambar 1.</b> Tampilan Aplikasi .....	1
<b>Gambar 2.</b> Panduan Input.....	4
<b>Gambar 3.</b> Panduan Tombol Run Topic Detection .....	4
<b>Gambar 4.</b> Panduan Selesai .....	5
<b>Gambar 5.</b> Import Google Translator dari deep_translator .....	6
<b>Gambar 6.</b> Contoh Penggunaan Dictionary.....	7
<b>Gambar 7.</b> File Excel Kata-kata Typo.....	7
<b>Gambar 8.</b> Alur Perbaikan Kata Typo.....	8
<b>Gambar 9.</b> Alur Translate Data .....	9
<b>Gambar 10.</b> Alur Pemisahan Row pada Translate Data .....	10
<b>Gambar 11.</b> Alur Ekstraksi Kata Kunci.....	11
<b>Gambar 12.</b> Alur Kategorisasi.....	13
<b>Gambar 13.</b> Alur set Stability.....	14
<b>Gambar 14.</b> Alur Finalisasi Kategori .....	15
<b>Gambar 15.</b> Alur Kerja Name Entity Recognition .....	15
<b>Gambar 16.</b> Hasil Name Entity Recognition.....	16
<b>Gambar 17.</b> Error modul Spacy.....	16
<b>Gambar 18.</b> File hasil .....	16

## A. Tentang Aplikasi

**Aplikasi Topic Detection** adalah sebuah aplikasi berbasis Excel VBA yang digunakan untuk mengolah sentimen atau komentar-komentar dari pelanggan terhadap sebuah produk ataupun sebuah bisnis. Beberapa proses utama pada aplikasi ini yaitu melakukan *Translate* dari komentar berbahasa Indonesia ke komentar berbahasa Inggris, melakukan Ekstraksi kata kunci dari sebuah komentar, melakukan kategorisasi komentar, melakukan ekstraksi lokasi dari komentar, terakhir melakukan *sentiment Analysis*.

## B. Tampilan Aplikasi

Di bawah ini merupakan tampilan awal dari Aplikasi Topic Detection Berbasis Excel VBA



Gambar 1. Tampilan Aplikasi

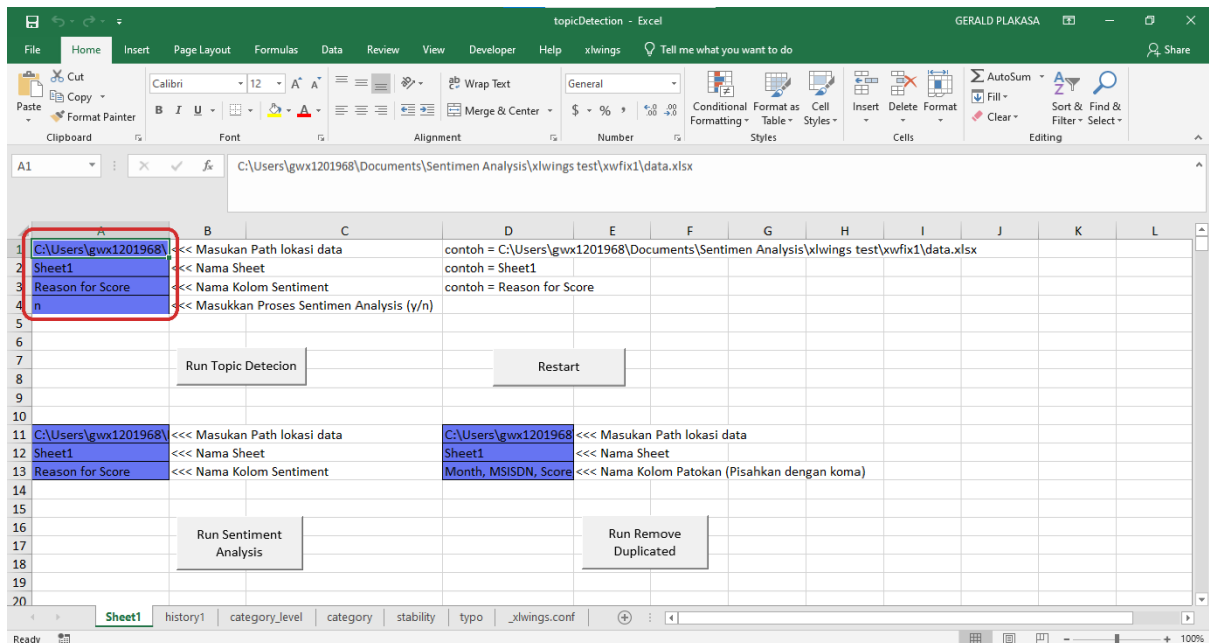
**Tabel 1.** Komponen dalam Aplikasi

No.	Nama Komponen	Deskripsi
1.	<i>Input Path Lokasi Data</i>	Merupakan tempat memasukkan <i>path</i> lokasi data disimpan, sebagai contohnya pada <i>path</i> “C:\Users\gwx1201968\Documents\Sentimen Analysis\xlwings test\xwfix1\data.xlsx”. Warna awal sebelum di <i>running</i> ialah berwarna biru, namun setelah di <i>running</i> jika ada <i>error</i> akan berubah menjadi merah sebaliknya akan berwarna Hijau.
2.	<i>Input Nama Sheet</i>	Merupakan tempat memasukkan nama dari <i>Sheet</i> pada data yang di simpan. Secara <i>Default</i> nama <i>Sheet</i> nya adalah Sheet1. Warna awal sebelum di <i>running</i> ialah berwarna biru, namun setelah di <i>running</i> jika ada <i>error</i> akan berubah menjadi merah sebaliknya akan berwarna Hijau.
3.	<i>Input Nama Kolom Sentimen</i>	Merupakan tempat memasukkan nama kolom sentimen / komentar yang di simpan. Secara <i>default</i> namanya ialah “ <i>Reason for Score</i> ”. Warna awal sebelum di <i>running</i> ialah berwarna biru, namun setelah di <i>running</i> jika ada <i>error</i> akan berubah menjadi merah sebaliknya akan berwarna Hijau.
4.	<i>Input Proses Sentimen Analisis</i>	Merupakan tempat memasukkan y/n untuk melakukan proses sentimen atau tidak.
5.	Informasi Proses	Ketika proses di <i>Running</i> dan data berhasil terbaca akan menampilkan informasi sejauh mana proses berjalan, 25%, 50%, 70%, atau selesai.
6.	Tombol <i>Run Topic Detection</i>	Jika semua <i>input</i> telah dimasukkan, menekan tombol <i>Run Topic Detection</i> akan memulai prosesnya.
7.	Tombol <i>Restart</i>	Merupakan tombol jika ingin mengembalikan tampilan menjadi <i>default</i> .
8.	<i>Input Kolom Patokan</i>	Merupakan tempat memasukkan nama – nama kolom patokan saat nanti dilakukan <i>drop duplicated</i> . Warna awal sebelum di <i>running</i> ialah berwarna biru, namun setelah di <i>running</i> jika ada <i>error</i> akan berubah menjadi merah sebaliknya akan berwarna Hijau.

9.	Tombol <i>Run Sentimen analysis</i>	Jika semua <i>input</i> telah dimasukkan, menekan tombol <i>Run Sentimen analysis</i> akan memulai hanya untuk proses sentimen analisis.
10.	Tombol <i>Run Remove Duplicated</i>	Jika semua <i>input</i> telah dimasukkan, menekan tombol <i>Run Remove Duplicated</i> akan memulai proses penghapusan data yang sama berdasarkan kolom patokannya.
11.	<i>Sheet "Sheet1"</i>	Merupakan <i>Sheet</i> utama dari Aplikasi
12.	<i>Sheet "history1"</i>	Merupakan <i>Sheet</i> untuk melihat <i>history</i> dari semua proses yang telah dilakukan
13.	<i>Sheet "category"</i>	Merupakan <i>Sheet</i> untuk mengatur kamus kata kunci kategori
14.	<i>Sheet "category_level"</i>	Merupakan <i>Sheet</i> untuk mengatur tingkat kepentingan pada masing-masing kategori
15.	<i>Sheet "stability"</i>	Merupakan <i>Sheet</i> untuk mengatur kamus kata kunci <i>Stability</i>
16.	<i>Sheet "typo"</i>	Merupakan <i>Sheet</i> untuk mengatur kata-kata <i>typo</i> dan kata yang benar
17.	<i>Sheet "_xlwings.conf"</i>	Merupakan <i>Sheet</i> untuk melakukan pengaturan aplikasi berdasarkan <i>library xlwings</i> sesuai dengan keadaan perangkat. Untuk menerapkannya <i>rename</i> nama <i>Sheet</i> menjadi <i>"xlwings.conf"</i>

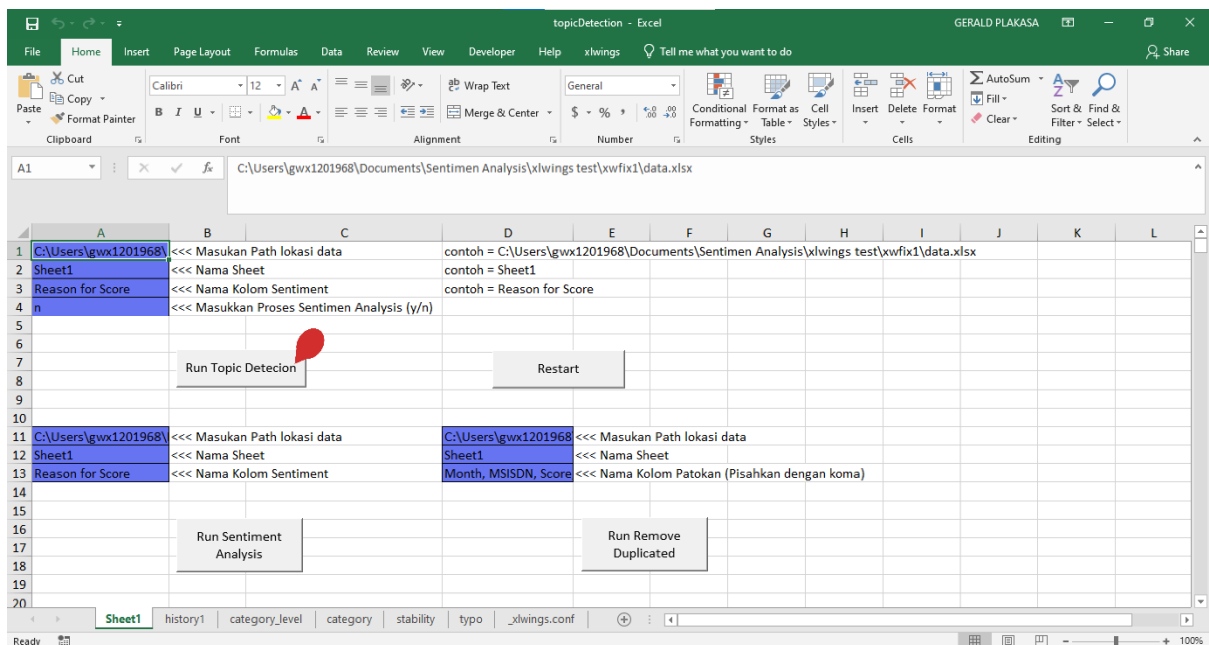
## C. Panduan Penggunaan Aplikasi

### 1. Masukkan *Input* yang dibutuhkan sesuai dengan ketentuan dan kebutuhan



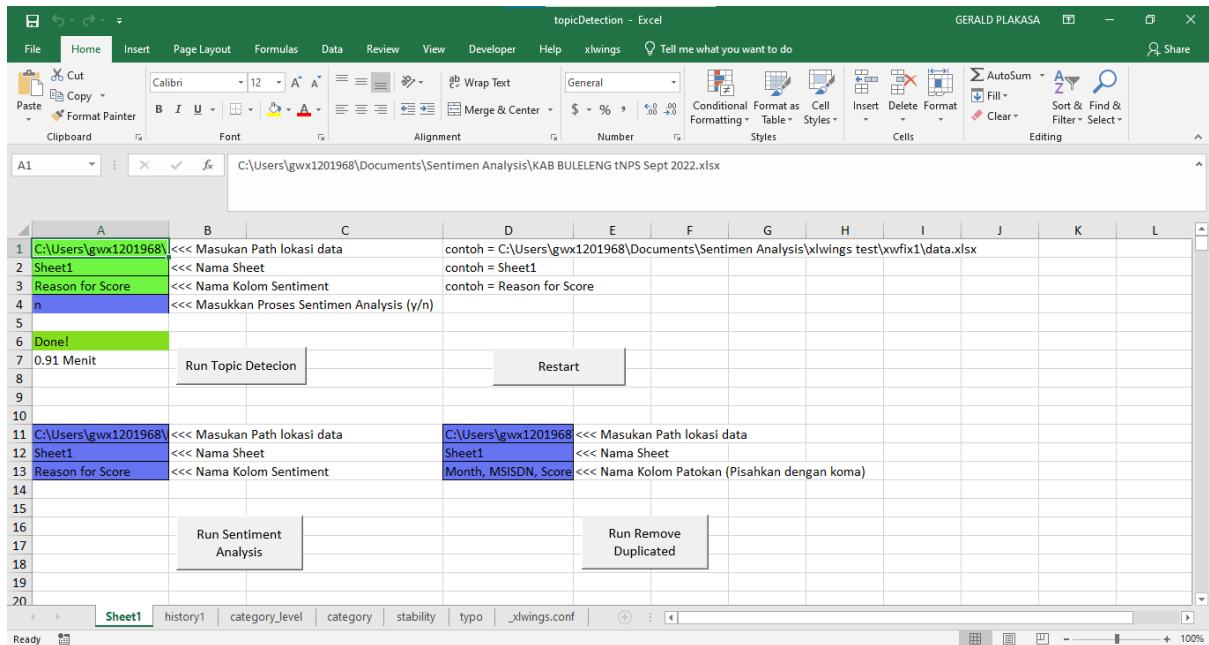
**Gambar 2.** Panduan Input

### 2. Tekan Tombol *Run Topic Detection*



**Gambar 3.** Panduan Tombol *Run Topic Detection*

3. Jika proses berhasil dan selesai akan tampil seperti di bawah ini



Gambar 4. Panduan Selesai

## D. Detail Proses Aplikasi

### 1. Data Preparation

Untuk mengolah komentar-komentar atau sentimen dari *customer* terhadap sebuah bisnis dapat menggunakan *Natural Language Processing* (NLP). NLP adalah kemampuan program komputer untuk memahami bahasa manusia seperti yang diucapkan dan ditulis atau disebut sebagai bahasa alami. Pada kasus saat ini, akan dilakukan deteksi topik dari komentar-komentar tersebut. Beberapa hal yang perlu dilakukan yaitu, *Translate* Bahasa dari Indonesia ke Inggris, Ekstraksi Kata Kunci dari Bahasa, Kategorisasi, Ekstraksi Lokasi, dan terakhir analisis Sentimen.

#### 1.1. Library Pendukung

Sebelum memulai, terdapat beberapa *library* pendukung yang digunakan untuk proses data dan teks. Beberapa *library* tersebut yaitu *numpy*, *pandas*, dan *nlTK*.

##### a. Numpy

*Numpy* merupakan sebuah *library Python* yang dapat membantu melakukan komputasi numerik ataupun statistik.

## b. Pandas

Pandas merupakan sebuah *library Python* yang membantu untuk mengolah data, dapat menampilkan, perubahan, menghapus, dan membuat data.

## c. nltk

nltk merupakan *Library* yang membantu proses pengolahan Bahasa alami

Untuk *Library – Library* yang membantu proses pada modul ini, semua telah tersedia pada dokumen **requirement.txt** , untuk instalasinya ketikkan “pip install -r requirement.txt”.

### 1.2. Read Data

Pertama hal yang dilakukan yaitu bagaimana caranya untuk mengambil data yang awalnya berbentuk Excel atau CSV ke dalam *notebook*. Dalam kasus ini dapat menggunakan Library Pandas. Pada Pandas terdapat fungsi khusus untuk mengambil data Excel atau CSV yaitu “read\_excel” atau “read\_csv”.

memanggil pd untuk read\_excel untuk data “data.xlsx” dan *Sheet* “Sheet1” lalu di simpan di variabel bernama df. Saat di panggil df di bawahnya akan muncul tabel Excel yang di baca. Lalu karena hanya ingin proses data sentimennya saja, maka ketik seperti bawah untuk mengambilnya

artinya mengambil kolom “Reason for Score” lalu di masukkan ke variabel df\_teks. Agar df\_teks tetap dalam bentuk *Dataframe* (array 2 dimensi) maka pada df digunakan kurung siku 2x. Jika kurung siku hanya 1x akan berbentuk *Series* (array 1 dimensi).

## 2. Translate Text

Setelah data sudah siap, Hal pertama yang harus dilakukan adalah melakukan *Translate* dari Bahasa Indonesia ke inggris. karena proses teks dengan Bahasa inggris sudah lebih *powerfull* dibandingkan dengan teks berbahasa Indonesia maka harus di ubah bahasanya. Untuk melakukannya diperlukan *library* tambahan untuk akses *Google Translator*. *Library*nya yaitu deep\_translator.

Dikarenakan deep\_translator tidak hanya *support* untuk *translate* menggunakan *Google Translator*, agar lebih spesifik ketikkan seperti di bawah ini.

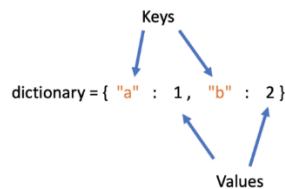
```
In [13]: from deep_translator import GoogleTranslator  
  
In [ ]: |
```

**Gambar 5.** Import Google Translator dari deep\_translator



## 2.1. Typo Text

Sebelum melakukan *translate* data perlu dilakukan pembenaran pada kata *typo* yang tidak bisa di *translate*. untuk melakukannya perlu di catat terlebih dahulu kata-kata yang secara *heuristic typo* dan kata yang benarnya menggunakan *dictionary*. *Dictionary* merupakan tipe penyimpanan data yang memiliki struktur seperti kamus, *Key-Value*.



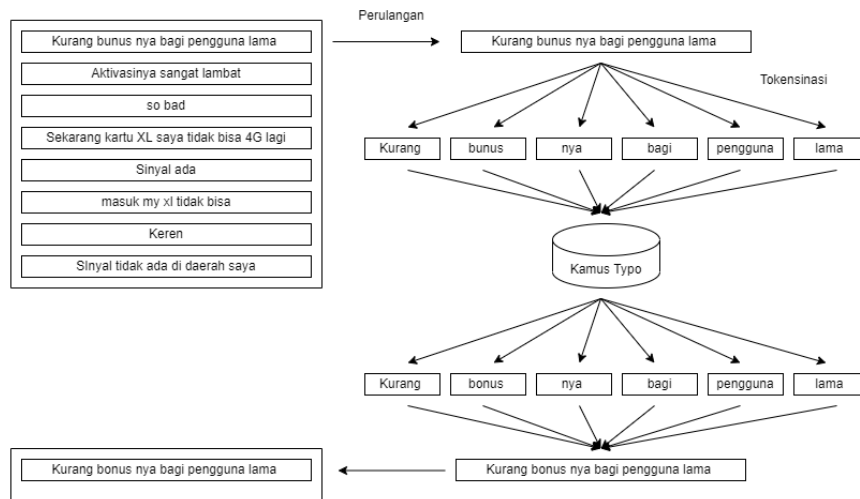
**Gambar 6.** Contoh Penggunaan *Dictionary*

Pada variabel kata *typo*, saya set *Key* merupakan kata salahnya dan *value* merupakan kata yang benarnya. Untuk datanya disimpan dalam *file* Excel seperti di bawah ini.

typo	benar
berinternet	akses internet
apiksix	aplikasi
aflikasi	aplikasi
apkl	aplikasi
apinya	aplikasinya
exiz	axis
axiz	axis
exis	axis
eksis	axis
Axiss	axis
axisnet	axis
exisnet	axis
xsis	axis
exsis	axis
bagud	bagus
byk	banyak
trabel	bermasalah
bonus	bonus
bobus	bonus
bunos	bonus

**Gambar 7.** File Excel Kata-kata *Typo*

Selanjutnya, kata *typo* yang ada pada data, akan di deteksi dan di *replace* dengan kata yang benarnya. Alurnya seperti di bawah ini.

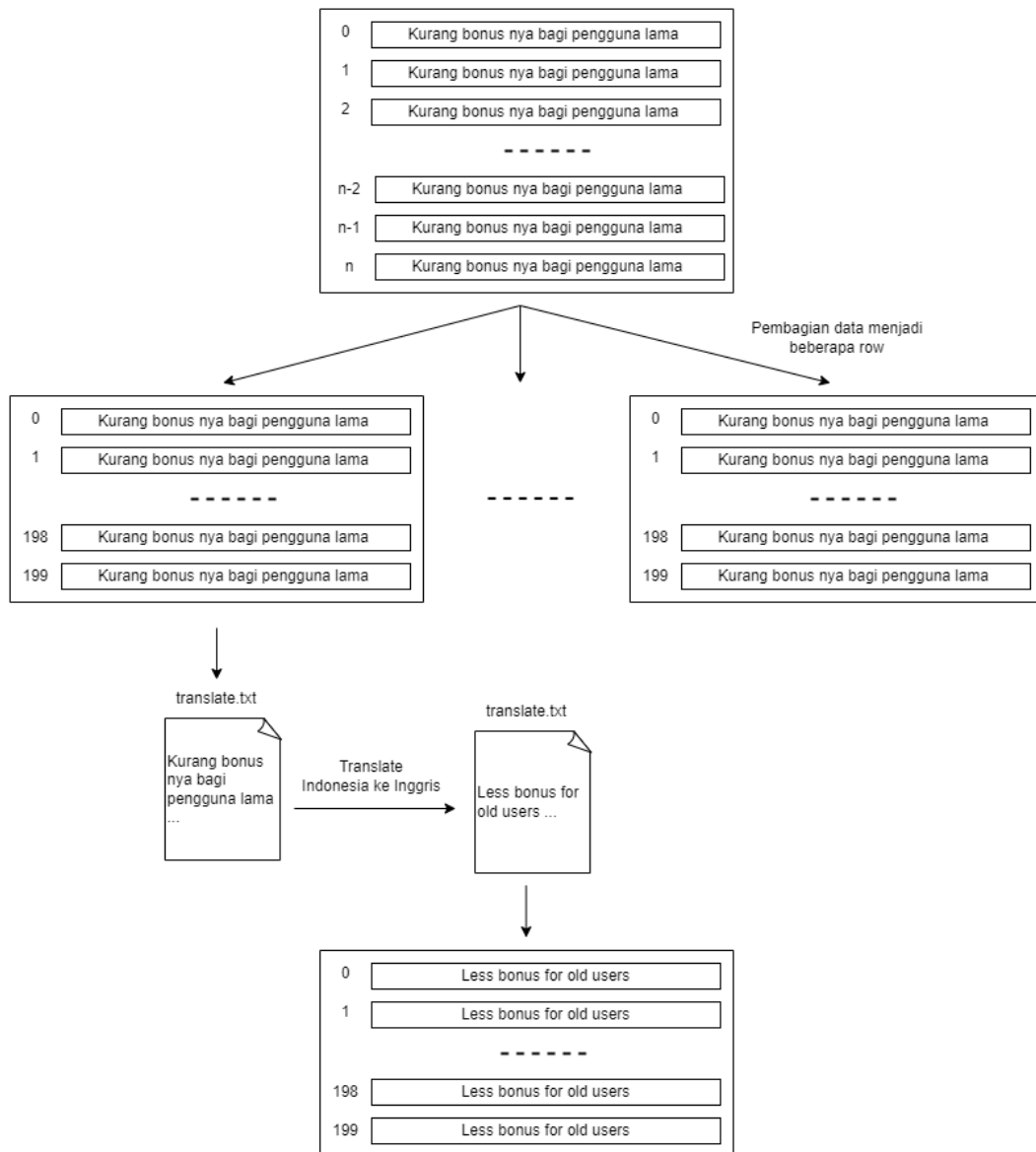


**Gambar 8.** Alur Perbaikan Kata *Typo*

Selain kata *typo* nya dibenarkan, yang diperbaiki juga pada bentuk katanya, seperti jika ada titik yang berdempetan antara 2 kata contoh “sinyal.buruk” agar saat *translate* datanya baik, harus diubah menjadi “sinyal. buruk”. Begitu pula jika ada kata seperti “sinyal . buruk” akan lebih baik *ditranslate* saat katanya menjadi “sinyal. buruk”. Terakhir yang saya lakukan yaitu jika ada karakter yang berulang lebih dari 2 kali akan di hapus menjadi 2 contohnya seperti “mahalllllll” akan menjadi “mahall” agar proses *translatenya* jadi baik.

## 2.2. *Translate Text*

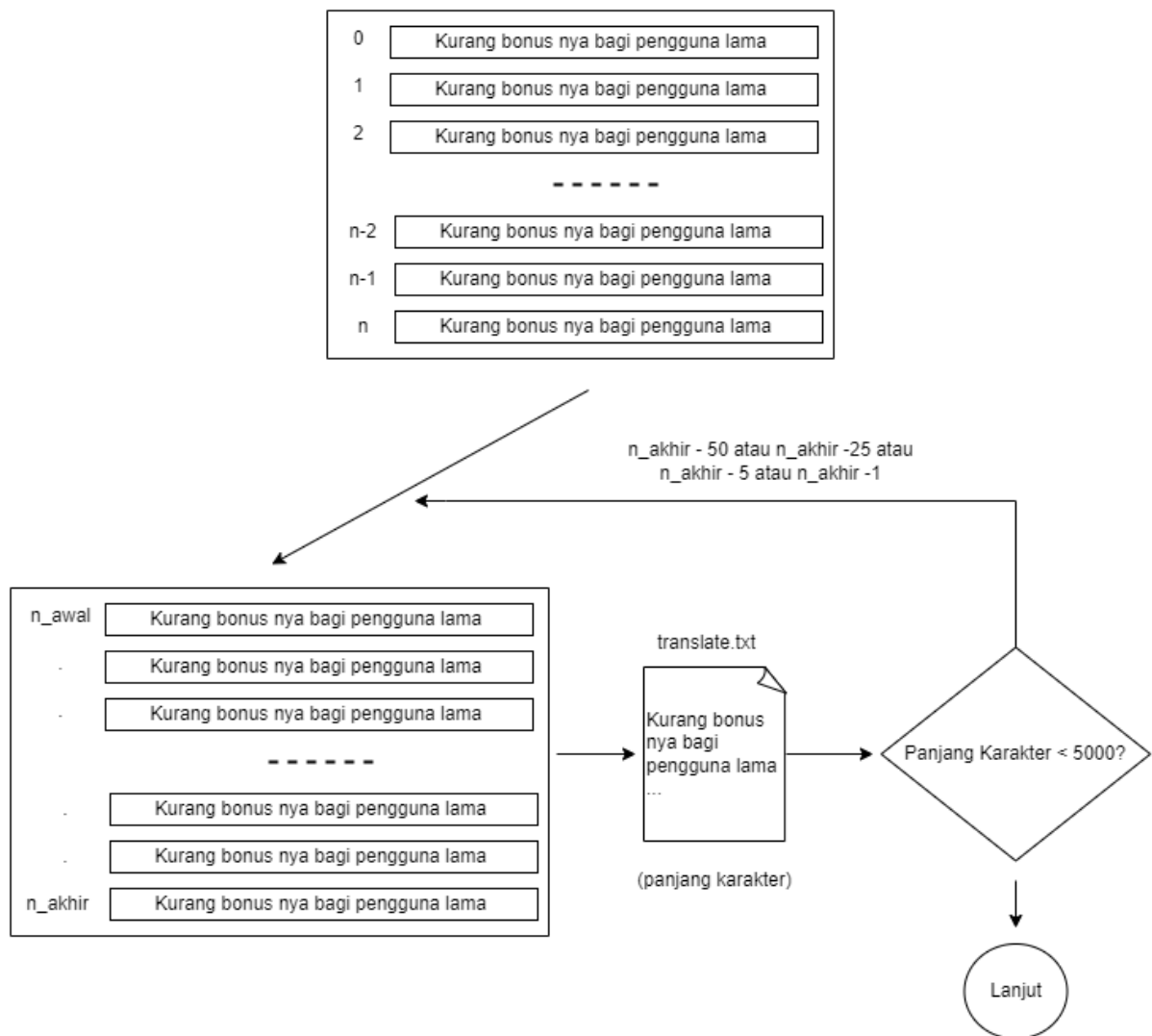
Selanjutnya melakukan *Translate* satu-satu dengan memasukkannya ke *file txt* lalu di *translate* seperti alur di bawah ini



**Gambar 9.** Alur *Translate Data*

Dikarenakan ketika *Translate* menggunakan *Google Translator*, memiliki batas maksimum karakter yaitu 5000 karakter. Untuk mengatasinya, dilakukan *Translate* setiap 200 baris pada data (dengan anggapan 200 baris < 5000 karakter). Tetapi juga untuk 200 baris tidak memiliki Panjang karakter yang pasti, jadi jika saat 200 baris jumlah karakternya lebih dari 5000, maka dijadikan 50 baris, jika masih lebih juga dijadikan 25, jika masih lebih juga dijadikan 5, dan jika masih lebih hanya 1 baris yang di *Translate*, dan untuk selanjutnya balik ke 200 baris lagi.

Selanjutnya di bawah ini alur untuk proses bagaimana cara memisah *row* sehingga setiap perulangan Panjang karakternya < 5000.



**Gambar 10.** Alur Pemisahan Row pada *Translate Data*

### 2.3. Lower

Selanjutnya karena kata “Python” dan “python” dikatakan berbeda oleh komputer, maka perlu di samakan semua teks dengan melakukan lower(). Teks di ambil untuk setiap hasil *Translate* dari kolom Translate2 lalu di jadikan huruf kecil semua, setelah itu hasilnya di simpan pada kolom *Translate*.

Pada kolom Translate2 (tanpa di lower) akan digunakan untuk proses NER atau ekstraksi lokasi sedangkan pada kolom *Translate* digunakan untuk proses lainnya maka keduanya harus dipisahkan di kolom yang berbeda.

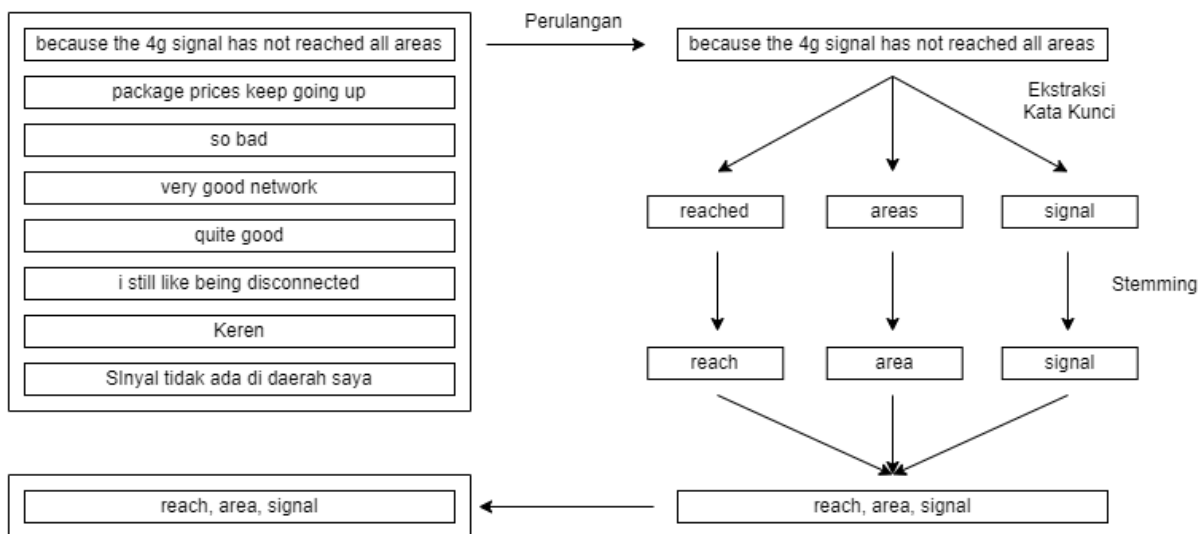
### 3. Keyword Extraction

Selanjutnya setelah semua teks di ubah menjadi Bahasa Inggris, tahapan berikutnya yaitu mengambil beberapa kata kunci yang mewakili teks tersebut.

`nlTK.stem` merupakan fungsi bantuan untuk melakukan *Stemming*. *Stemming* adalah proses penghapusan kata depan atau kata belakang, menjadi kata aslinya. Contohnya kata “planning” dan kata “plans” jika dilakukan *Stemming* hasilnya akan sama-sama menjadi “plan”.

Melakukan *Stemming* menggunakan *PorterStemmer*, pertama membuat *object* yang di simpan pada variabel `ps`. Beberapa parameter tambahannya juga di atur untuk Bahasa, maksimal Panjang suku kata, *Threshold* penghapusan duplikat, dan jumlah kata kunci yang ingin di hasilkan. Lalu untuk ekstraksi kata kuncinya juga dibuatkan *object* yang disimpan pada `kw_extractor`.

Setelah di *setting*, berikut ini alur untuk proses ekstraksi kata kunci nya pada data



**Gambar 11.** Alur Ekstraksi Kata Kunci

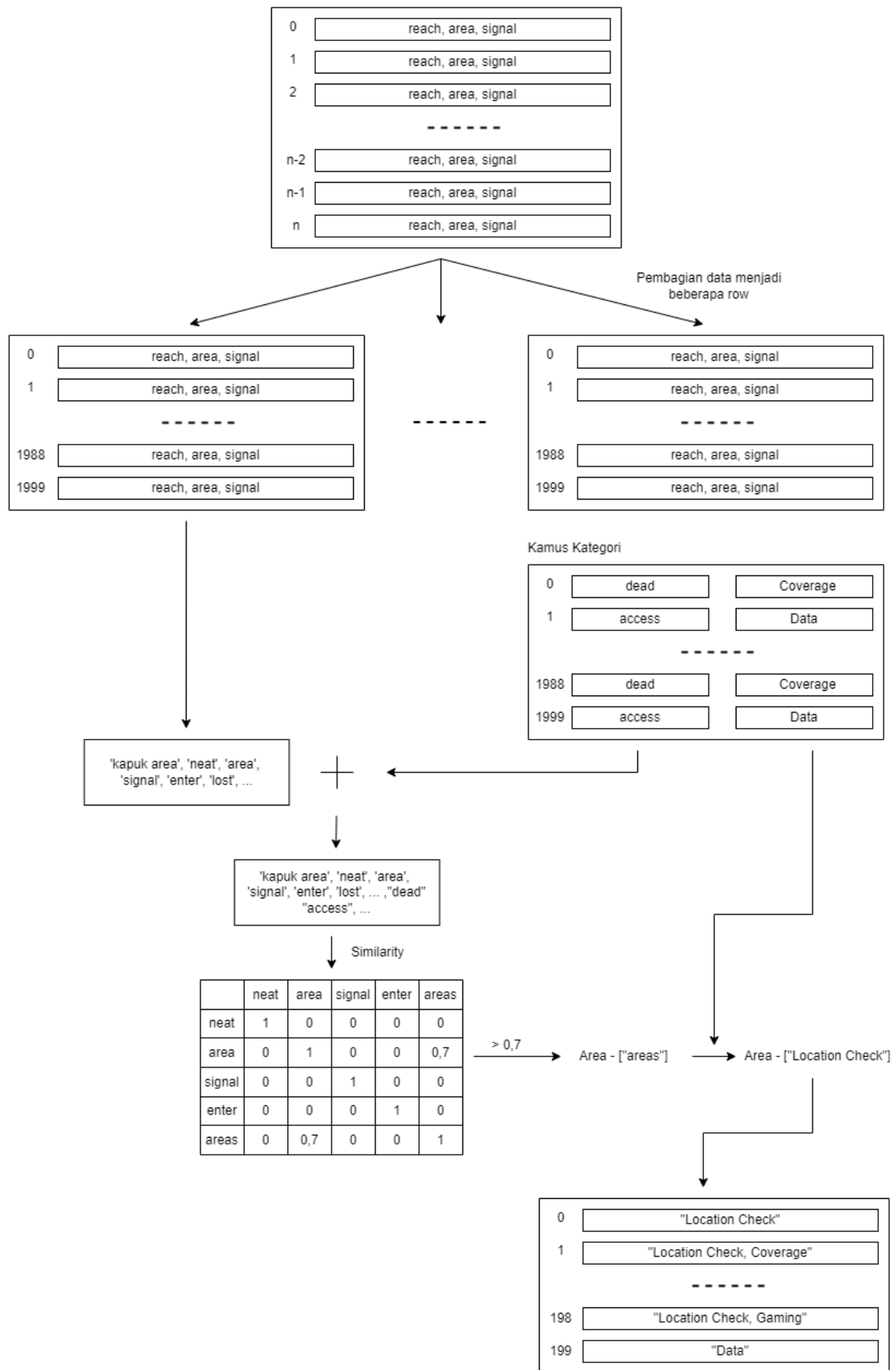
## **4. Kategorisasi**

### **4.1. Persiapan Data**

Sebelum mulai mengategorisasikan kata kunci yang sudah di ekstrak sebelumnya, terlebih dahulu mengambil data kategori yang sudah disiapkan. Terdapat *file* Category.xlsx yang berisikan kumpulan kata dan kategorinya, digunakan untuk proses selanjutnya.

### **4.2. Categorization**

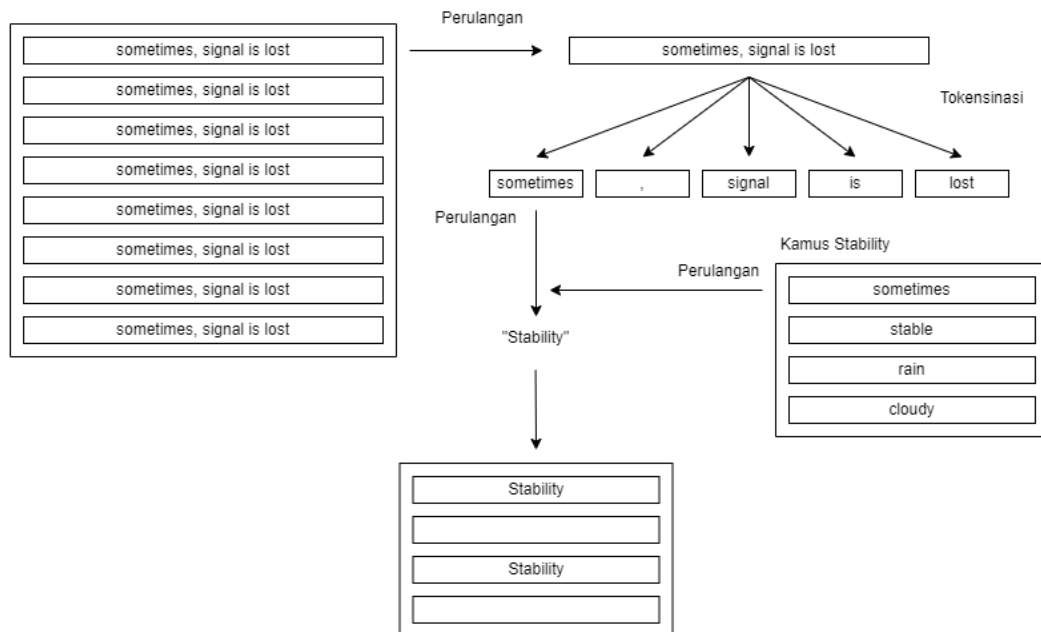
Selanjutnya buat sebuah fungsi yang berguna untuk membersihkan kata dari spasi berlebih, dari karakter spesial, dan dari *stopword*. *Stopword* merupakan kata umum yang mempunyai fungsi tapi tidak mempunyai arti. Karena tidak memiliki arti maka harus dibersihkan. *Punctuation* merupakan karakter spesial seperti (“,” “-“, “[“,”]”, dll.) karakter tersebut dapat dihapuskan karena tidak memiliki arti. Strip merupakan penghapusan spasi yang berlebih di awal kata ataupun di akhir kata. Selanjutnya melakukan kategorisasi mengikuti alur berikut ini.



**Gambar 12.** Alur Kategorisasi

### 4.3. Set Stability

Selanjutnya dapat set *Stability* dari kamus kata dan kategori *Stability* nya, Untuk proses ini sedikit berbeda dengan kategorisasi karena terdapat beberapa kata yang biasanya tidak masuk ke dalam kata kunci seperti “*sometimes*”, dll. Maka dari itu untuk proses *Stability* memakan waktu lebih lama karena akan di cek teksnya satu – satu setiap *row*. Lalu untuk alurnya seperti di bawah ini

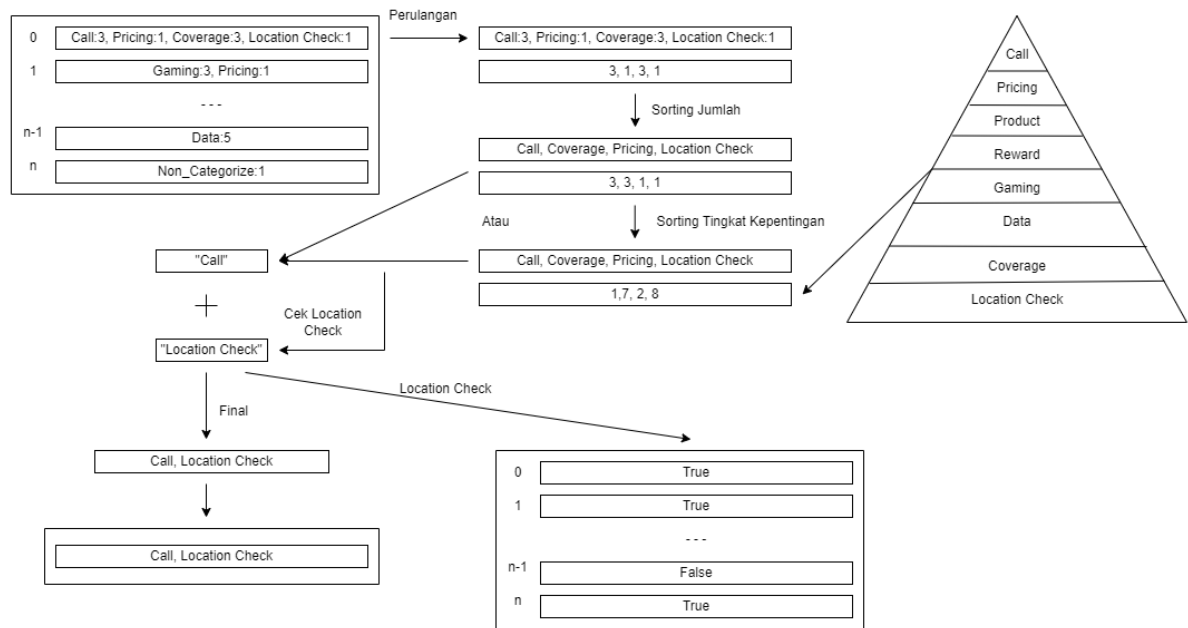


**Gambar 13.** Alur set Stability

### 4.4. Finalisasi Kategori

Terakhir setelah didapatkan kategori untuk masing-masing teks, maka selanjutnya mengambil kategori finalnya berdasarkan tingkat kepentingan kategori. Selanjutnya buat fungsi yang berguna untuk melakukan pengurutan 2 *list* berdasarkan pada nilai *list* pertama. Alur finalisasi kategori yaitu seperti di bawah ini.



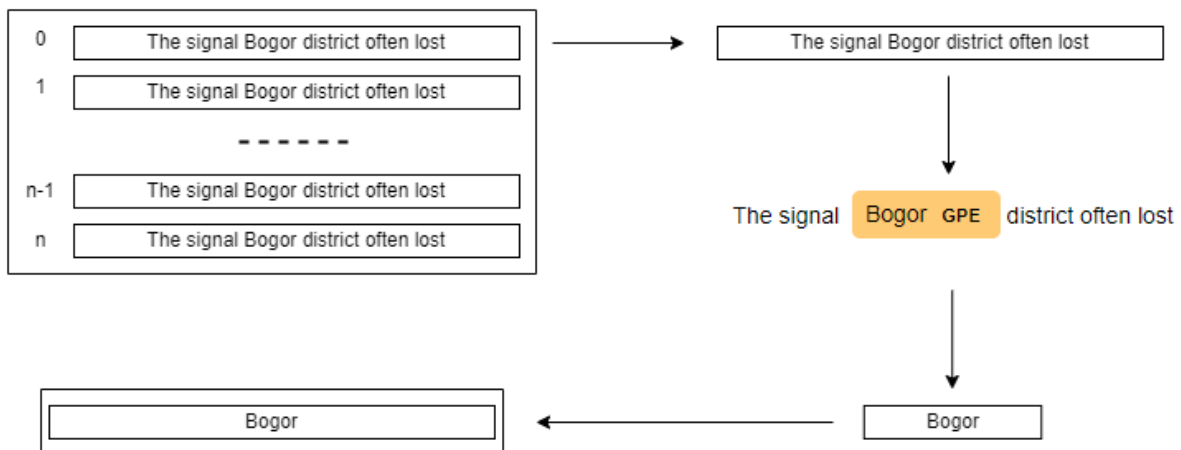


**Gambar 14.** Alur Finalisasi Kategori

Selanjutnya jika ada *Non\_network* yang termasuk juga *Location Check*, maka harus di hapuskan. lalu simpan hasil akhirnya dan juga hasil label *Location Check*. Jika *Stability* terdapat di dalam *Non\_Network* dan *Non\_Categorize* harus dihapuskan juga. Lalu, terakhir tinggal memasukkan *Stability* di kategori final.

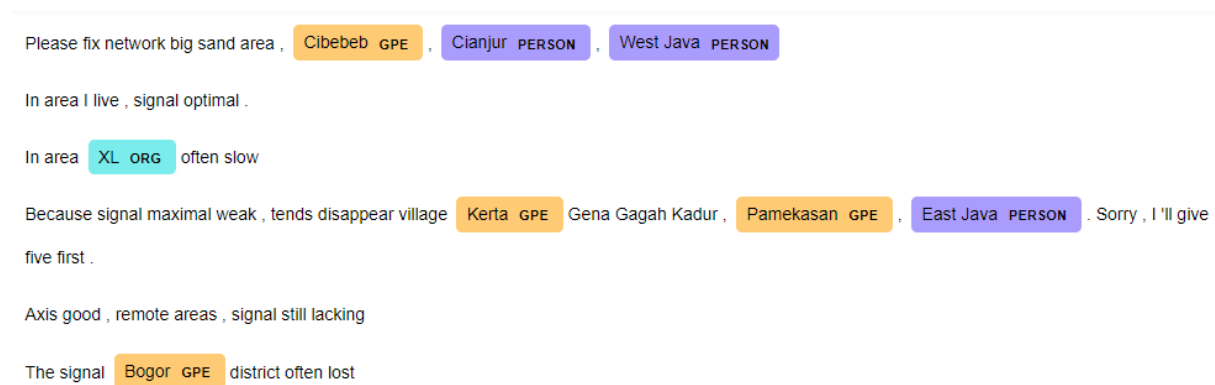
## 5. Name Entity Recognition

Setelah mendapatkan teks yang mengandung *Location Check* dan kategori di dapatkan, maka selanjutnya itu mengambil entitas lokasi dari teks tersebut. Untuk melakukan hal tersebut dinamakan NER (*Name Entity Recognition*) menggunakan *library* Spacy. Alur kerja dari *Name Entity Recognition* sebagai berikut.



**Gambar 15.** Alur Kerja Name Entity Recognition

Hasilnya akan seperti di bawah ini



**Gambar 16.** Hasil *Name Entity Recognition*

Jika saat eksekusi modulnya belum ada, dapat melakukan *install* menggunakan “pip install” nama modul. Apabila terdapat eror seperti di bawah ini

```
-----  
ModuleNotFoundError                                Traceback (most recent call last)  
Input In [36], in <cell line: 4>()  
      2 from spacy import displacy  
      3 from collections import Counter  
----> 4 import en_core_web_sm  
      5 nlp = en_core_web_sm.load()  
      7 location = []  
  
ModuleNotFoundError: No module named 'en_core_web_sm'
```

**Gambar 17.** *Error modul Spacy*

Untuk instalasinya dapat ketikan “python -m spacy download en\_core\_web\_sm” pada *Powershell* nya. Lalu dapat mengeksekusi kode di atas.

## 6. *Post Processing*

Terakhir setelah selesai semua terkait prosesnya, terakhir tinggal merapikan tabel dan menyimpannya kembali ke dalam *file* Excel.

Jika selesai simpan ke aalam bentuk excel dan akan muncul *file* tersebut di folder yang sama dengan *file notebook*



**Gambar 18.** *File hasil*