

Hasil :

Kalimat Awal

Menko Polhukam Mahfud Md menyatakan pengelolaan dana otonomi khusus (otsus) Papua tidak beres. Karenanya, dana otsus dinaikkan menjadi 2,25 persen dari Dana Alokasi Khusus APBN

Tokenizing

Mengecilkan dan menghilangkan karakter khusus

string(171) "menko polhukam mahfud md menyatakan pengelolaan dana otonomi khusus otsus papua tidak beres karenanya dana otsus dinaikkan menjadi 225 persen dari dana alokasi khusus apbn"

Menghilangkan angka

string(168) "menko polhukam mahfud md menyatakan pengelolaan dana otonomi khusus otsus papua tidak beres karenanya dana otsus dinaikkan menjadi persen dari dana alokasi khusus apbn"

Hasil akhir Tokenisasi

No	Token
1	menko
2	polhukam
3	mahfud
4	md
5	menyatakan
6	pengelolaan
7	dana
8	otonomi
9	khusus
10	otsus
11	papua
12	tidak
13	beres
14	karenanya
15	dana
16	otsus
17	dinaikkan
18	menjadi
19	persen
20	dari
21	dana
22	alokasi
23	khusus
24	apbn

Filtering

Tala Stopword

No	Kata
1	ada
2	adalah
3	adanya
4	adapun
5	agak
6	agakny
7	agar
8	akan
9	akankah
10	akhir
11	akhiri
12	akhirnya
13	aku
14	akulah
15	amat
16	amatlah
17	anda
18	andalah
19	antar
20	antara
21	antaranya
22	apa

730	tinggi
731	toh
732	tunjuk
733	turut
734	tutur
735	tuturnya
736	ucap
737	ucapnya
738	ujar
739	ujarnya
740	umum
741	umumnya
742	ungkap
743	ungkapnya
744	untuk
745	usah
746	usai
747	waduh
748	wah
749	wahai
750	waktu
751	waktunya
752	walau
753	walaupun
754	wong
755	yaitu
756	yakin
757	yakni
758	yang

Hasil Filtering

No	Kata
1	menko
2	polhukam
3	mahfud
4	md
5	pengelolaan
6	dana
7	otonomi
8	khusus
9	otsus
10	papua
11	beres
12	dana
13	otsus
14	dinaikkan
15	persen
16	dana
17	alokasi
18	khusus
19	apbn

Source Code

```
<?php
include "function_tala_stopword.php";
$kalimat_awal="Menko Polhukam Mahfud Md menyatakan pengelolaan dana otonomi khusus (otsus) Papua
tidak beres. Karenanya, dana otsus dinaikkan menjadi 2,25 persen dari Dana Alokasi Khusus APBN";
echo "<h1> Kalimat Awal </h1>";
echo $kalimat_awal;
echo "<br><br>";
$kalimat=strtolower($kalimat_awal);
$kalimat_remove_karakter=preg_replace("/[^A-Za-z0-9\ ]/", "", $kalimat);
echo "<h1> Tokenizing </h1>";
echo "<h4> Mengecilkkan dan menghilangkan karakter khusus </h4>";
var_dump($kalimat_remove_karakter);
echo "<br><br>";

$kalimat_remove_angka=preg_replace('/\d+/u', '', $kalimat_remove_karakter);
echo "<h4> Menghilangkan angka </h4>";
var_dump($kalimat_remove_angka);
echo "<br><br>";
$token_kalimat=explode(' ', $kalimat_remove_angka);
$token_kalimat2=array_filter($token_kalimat, function($value) { return !is_null($value) && $value
!= ' '; });

// var_dump($kalimat);
// display $token_kalimat as table
echo "<h4> Hasil akhir Tokenisasi</h4>";
echo "<table border='1'>";
echo "<tr><th> No </th><th> Token </th></tr>";
$no=1;
foreach ($token_kalimat2 as $token) {
    echo "<tr><td>".$no."</td><td>".$token."</td></tr>";
    $no++;
}
echo "</table>";

// var_dump($token_kalimat);
echo"<br><br>";

echo "<h1>Filtering</h1>";
$tala_word=tala_stopword();
?>
<h4> Tala Stopword </h4>
<table border='1'>
<tbody style="height: 100px; overflow-y: scroll;">
<tr><th> No </th><th> Kata </th></tr>
<?php
$no=1;
foreach ($tala_word as $tw) {
    echo "<tr><td>".$no."</td><td>".$tw."</td></tr>";
    $no++;
}
?>
</tbody>
</table>
<br><br>
<?php
$kalimat_filtering=array_diff($token_kalimat, $tala_word);
$kalimat_filtering2=array_filter($kalimat_filtering, function($value) { return !is_null($value) &&
$value != ' '; });
echo "<h4> Hasil Filtering </h4>";
// var_dump($kalimat_filtering);
// display $kalimat_filtering as table
echo "<table border='1'>";
echo "<tr><th> No </th><th> Kata </th></tr>";
$no=1;
foreach ($kalimat_filtering2 as $kf) {
    echo "<tr><td>".$no."</td><td>".$kf."</td></tr>";
    $no++;
}
echo "</table>";

// stemming process from $kalimat_filtering2
echo "<br><br>";
echo "<h1> Stemming </h1>";
?>
```

```

<?php
function tala_stopword(){
    $tala_stopword="ada adalah adanya adapun agak agaknya agar akan akankah akhir akhiri akhirnya
aku akulah amat amatlah anda andalah antar antara antaranya apa apaan apabila apakah apalagi
apatah artinya asal asalkan atas atau atukah ataupun awal awalnya bagi bagaikan bagaimana
bagaimanakah bagaimanapun bagi bagian bahkan bahwa bahwasanya baik bakal bakalan balik banyak
bapak baru bawah beberapa begini beginian beginilah begitu begitukah begitulah begitupun
bekerja belakang belakangan belum belumlah benar benarkah benarlah berada berakhir berakhirlah
berakhirnya berapa berapakah berapalah berapapun berarti berawal berbagai berdatangan beri berikan
berikut berikutnya berjumlah berkali-kali berkata berkehendak berkeinginan berkenaan berlainan
berlalu berlangsung berlebihan bermacam bermacam-macam bermaksud bermula bersama bersama-sama
bersiap bersiap-siap bertanya bertanya-tanya berturut berturut-turut bertutur berujar berupa besar
betul betulkah biasa biasanya bila bilakah bisa bisakah boleh bolehkah bolehlah buat bukan
bukankah bukanlah bukannya bulan bung cara caranya cukup cukupkah cukuplah cuma dahulu dalam dan
dapat dari daripada datang dekat demi demikian demikianlah dengan depan di dia diakhiri
diakhirinya dialah diantara diantaranya diberi diberikan diberikannya dibuat dibuatnya didapat
didatangkan digunakan diibaratkan diibaratkannya diingat ingatkan diinginkan dijawab dijelaskan
dijelaskannya dikarenakan dikatakan dikatakannya dikerjakan diketahui diketahuinya dikira
dilakukan dilalui dilihat dimaksud dimaksudkan dimaksudkannya dimaksudnya diminta dimintai
dimisalkan dimulai dimulailah dimulainya dimungkinkan dini dipastikan diperbuat diperbuatnya
dipergunakan diperkirakan diperlihatkan diperlukan diperlukannya dipersoalkan dipertanyakan
dipunyai diri dirinya disampaikan disebut disebutkan disebutkannya disini disinilah ditambahkan
ditandaskan ditanya ditanyai ditanyakan ditegaskan ditunjukan ditunjuk ditunjuki ditunjukkan
ditunjukkannya ditunjuknya dituturkan dituturkannya diucapkan diucapkannya diungkapkan dong dua
dulu empat enggak enggaknya entah entahlah guna gunakan hal hampir hanya hanyalah hari harus
haruslah harusnya hendak hendaklah hendaknya hingga ia ialah ibarat ibaratkan ibaratnya ibu ikut
ingat ingat-ingat ingin inginkah inginkan ini inikah inilah itu itukah itulah jadi jadilah jadinya
jangan janganlah janganlah jauh jawab jawaban jawabnya jelas jelaskan jelaslah jelaslah jika
jikalau juga jumlah jumlahnya justru kala kalau kalaulah kalaupun kalian kami kamilah kamu kamulah
kan kapan kapankah kapanpun karena karenanya kasus kata katakan katakanlah katanya ke keadaan
kebetulan kecil kedua keduanya keinginan kelamaan kelihatan kelihatannya kelima keluar kembali
kemudian kemungkinan kemungkinannya kenapa kepada kepadanya kesampaian keseluruhan keseluruhannya
keterlaluan ketika khususnya kini kinilah kira kira-kira kiranya kita kitalah kok kurang lagi
lagian lah lain lainnya lalu lama lamanya lanjut lanjutnya lebih lewat lima luar macam maka
makanya makin malah malahan mampu mampukah mana manakala manalagi masa masalah masalahnya masih
masihkah masing masing-masing mau maupun melainkan melakukan melalui melihat melihatnya memang
memastikan memberi memberikan membuat memerlukan memihak meminta memintakan memisalkan memperbuat
mempergunakan memperkirakan memperlihatkan mempersiapkan mempersoalkan mempertanyakan mempunyai
memulai memungkinkan menaiki menambahkan menandakan menanti menanti-nanti menanti menanti
menanyai menanyakan mendapat mendapatkan mendatang mendatangi mendatangkan menegaskan mengakhiri
mengapa mengatakan mengatakannya mengenai mengerjakan mengetahui menggunakan menghendaki
mengibaratkan mengibaratkannya mengingat mengingatkan menginginkan mengira mengucapkan
mengucapkannya mengungkapkan menjadi menjawab menjelaskan menuju menunjuk menunjuki menunjukkan
menunjuknya menurut menuturkan menyampaikan menyangkut menyatakan menyebutkan menyeluruh
menyiapkan merasa mereka merekalah merupakan meski meskipun meyakini meyakinkan minta mirip misal
misalkan misalnya mula mulai mulailah mulanya mungkin mungkinkah nah naik namun nanti nantinya
nyaris nyatanya oleh olehnya pada padahal padanya pak paling panjang pantas para pasti pastilah
penting pentingnya per percuma perlu perlukah perlunya pernah persoalan pertama pertama-tama
pertanyaan pertanyakan pihak pihaknya pukul pula pun punya rasa rasanya rata rupanya saat saatnya
saja sajalah saling sama sama-sama sambil sampai sampai-sampai sampaikan sana sangat sangatlah
satu saya sayalah se sebab sebabnya sebagai sebagaimana sebagainya sebagian sebaik sebaik-baiknya
sebaiknya sebaliknya sebanyak sebegini sebegini sebelum sebelumnya sebenarnya seberapa sebesar
sebetulnya sebisanya sebuah sebut sebutlah sebutnya secara secukupnya sedang sedangkan sedemikian
sedikit sedikitnya seenaknya segala segalanya segera seharusnya sehingga seingat sejak sejauh
sejenak sejumlah sekadar sekadarnya sekali sekali-kali sekalian sekaligus sekalipun sekarang
sekarang sekecil seketika sekiranya sekitar sekitarnya sekurang-kurangnya sekurangnya sela selain
selaku selalu selama selama-lamanya selamanya selanjutnya seluruh seluruhnya semacam semakin
semampu semampunya semasa semasih semata semata-mata semaunya sementara semisal semisalnya sempat
semua semuanya semula sendiri sendirian sendirinya seolah seolah-olah seorang sepanjang
sepentasnya septantasnyalah seperlunya seperti seperti halnya sering seringnya serta serupa
sesaat sesama sesampai sesegera sesekali seseorang sesuatu sesuatunya sesudah sesudahnya setelah
setempat setengah seterusnya setiap setiba setibanya setidak-tidaknya setidaknya setinggi sesuai
sewaktu siap siapa siapakah siapapun sini sinilah soal soalnya suatu sudah sudahkah sudahlah
supaya tadi tadinya tahu tahun tak tambah tambahannya tampak tampaknya tandas tandasnya tanpa tanya
tanyakan tanyanya tapi tegas tegasnya telah tempat tengah tentang tentu tentulah tentunya tepat
terakhir terasa terbanyak terdahulu terdapat terdiri terhadap terhadapnya teringat teringat-ingat
terjadi terjadilah terjadinya terkira terlalu terlebih terlihat termasuk ternyata tersampaikan
tersebut tersebutlah tertentu tertuju terus terutama tetap tetapi tiap tiba tiba-tiba tidak
tidakkah tidaklah tiga tinggi toh tunjuk turut tutur tuturnya ucap ucapannya ujar ujarannya umum
umumnya ungkap ungkapnya untuk usah usai waduh wah wahai waktu waktunya walau walaupun wong yaitu
yakini yakni yang";
    $tala_array=explode(" ", $tala_stopword);
    return $tala_array;
}

```